

QY	1	GGTCTGCACTGTATCCGAGCGGACAGATATCTGAGAACTCCACGGCAACGACAGATAC	60
Db	1	GGTCTGCACTGTATCCGAGCGGACAGATATCTGAGAACTCCACGGCAACGACAGATAC	60
QY	61	CGAGTACCAGTGTATCCAGACAGACATCCCCAGGAAAGCTACCGCTGTGGCCATC	120
Db	61	CGAGTACCAGTGTATCCAGACAGACATCCCCAGGAAAGCTACCGCTGTGGCCATC	120
QY	121	CTACCAACAAGGAGCTGCTCTTCACTGTTCAACTGAGCTGATGATCTG	180
Db	121	CTACCAACAAGGAGCTGCTCTTCACTGTTCAACTGAGCTGATGATCTG	180
QY	181	TGAGAGCCATGCCAGGTGCGGGCCCTTGTGTATCCAAACGACCACTGTGACAGTGA	240
Db	181	TGAGAGCCATGCCAGGTGCGGGCCCTTGTGTATCCAAACGACCACTGTGACAGTGA	240
QY	241	GCCAGTGGAGAAAGCCCTTCCAAAGGAGATGGACAGACCTCTGAGAGTTGATGATAG	300
Db	241	GCCAGTGGAGAAAGCCCTTCCAAAGGAGATGGACAGACCTCTCTGAGAGTTGATGATAG	300
QY	301	TGATCCCCCATCGAAGTCAAGAGGAGTGTAGAGTGAATGAGAGAGTATACCTGTCT	360
Db	301	TGATCCCCCATCGAAGTCAAGAGGAGTGTAGAGTGAATGAGAGAGTATACCTGTCT	360
QY	361	TCAAGGCAATGAATTTAGGAGGAATGTGCTTGCCTCCAGAAAGAGAAATTCACGCTGT	420
Db	361	TCAAGGCAATGAATTTAGGAGGAATGTGCTTGCCTCCAGAAAGAGAAATTCACGCTGT	420
QY	421	TTACTCTCACTCTGCCCCCAGGTGGGAGCTGTCTTTTCAAGA CTGATGAGACC	480
Db	421	TTACTCTCACTCTGCCCCCAGGTGGGAGCTGTCTTTTCAAGA CTGATGAGACC	480
QY	481	AAGTGTCCCTGATCCCAACAAGACAATATGTGAAGGCTCTGCTGACCTATCTGAG	540
Db	481	AAGTGTCCCTGATCCCAACAAGACAATATGTGAAGGCTCTGCTGACCTATCTGAG	540
QY	541	GGCTCGGCTGACAGCTGATCTATCTCAGACAGCTGGGCTTGCCTGTGGAGGAGTGA	600
Db	541	GGCTCGGCTGACAGCTGATCTATCTCAGACAGCTGGGCTTGCCTGTGGAGGAGTGA	600
QY	601	GCACTGGCAGACATGCATGTCACTGGGAAACCCCTGCAACAAGCTTAATCCACACA	660
Db	601	GCACTGGCAGACATGCATGTCACTGGGAAACCCCTGCAACAAGCTTAATCCACACA	660
QY	661	GACAGATGTGACAGAGCAAACTGTCAATATGCAATGTAAATGTGAGTTTACAG	720
Db	661	GACAGATGTGACAGAGCAAACTGTCAATATGCAATGTAAATGTGAGTTTACAG	720
QY	721	CTTAGCTATGAGGACTGTGCTGCTCTAGTCCAGAAATCATGGGGGTATGATCGCTCTCA	780
Db	721	CTTAGCTATGAGGACTGTGCTGCTCTAGTCCAGAAATCATGGGGGTATGATCGCTCTCA	780
QY	781	ACCGTGGGCTGTAAACCAAGCTCAAGCTATCTCCCATGGGGGCTGTGCCCCCTCT	840
Db	781	ACCGTGGGCTGTAAACCAAGCTCAAGCTATCTCCCATGGGGGCTGTGCCCCCTCT	840
QY	841	GGAGCGTTCCGTGGGAGCCCATCATCTGTTTCAATAGTGTAGAAATGTAGCTTAAAG	900
Db	841	GGAGCGTTCCGTGGGAGCCCATCATCTGTTTCAATAGTGTAGAAATGTAGCTTAAAG	900

Db 1381 AAAAAAAAAAGCGGTC 1397

ORGANISM    Homo sapiens

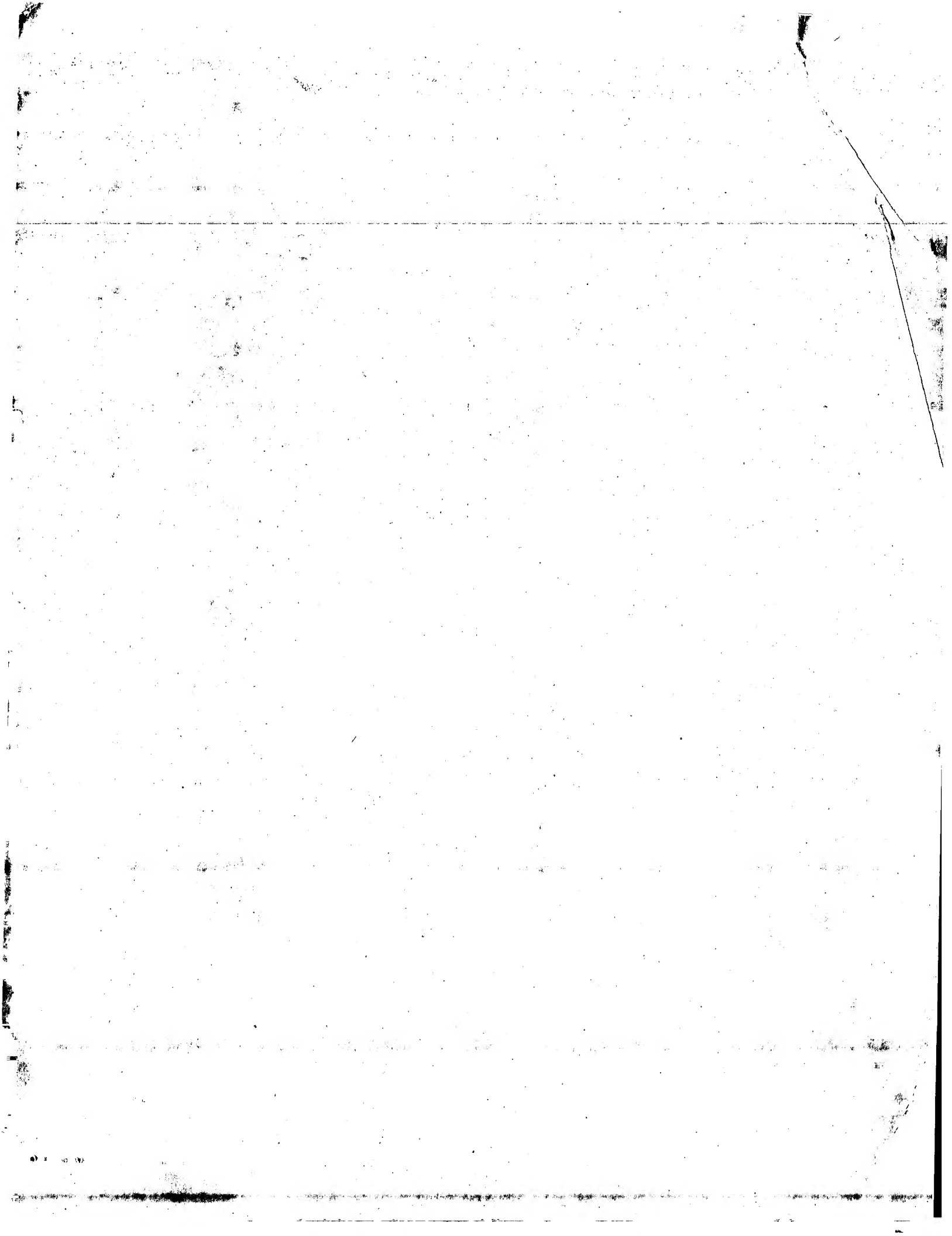
MEDLINE 99063792

REFERENCE 3 (bases 1 to 128466)

REFERENCE 4 (bases 1 to 128466) :

COMMENT On Nov 7, 2001 this sequ

Center: Washington University Genome Sequencing Center  
Center code: WUGSC





Web site: <http://genome.wustl.edu/gsc>  
 Contact: [sapiens@wustl.edu](mailto:sapiens@wustl.edu)  
 Summary Statistics  
 Center Project name: H\_NH0564016

NOTICE: This sequence may not represent the entire insert of this clone. It may be shorter because we only sequence overlapping clone sections once, or longer because we provide a small overlap between neighboring data submissions.

This sequence was finished as follows unless otherwise noted:  
 all regions were double stranded, sequenced with an alternate chemistry, or covered by high quality data (i.e., phred quality >= 30); an attempt was made to resolve all sequencing problems, such as compressions and repeats; all regions were covered by sequence from more than one subclone; and the assembly was confirmed by restriction digest.

MAPPING INFORMATION:  
 Mapping information for this clone was provided by Dr. John D. McPherson, Department of Genetics, Washington University, St. Louis MO. For additional information about the map position of this sequence, see <http://genome.wustl.edu/gsc>

SOURCE INFORMATION:  
 The RPCI-11 human BAC library was made from the blood of one male donor, as described by Osoegawa, K., Moon, P.Y., Zhao, B., Frengen, E., Tareno, M., Catanesi, J.J., and de Jong, P.O. (1998) An improved approach for construction of bacterial artificial chromosome libraries. Genomics 51:1-8. The clone may be obtained either from Research Genetics, Inc. (<http://www.resgen.com>) or Pieter de Jong and coworkers at the Roswell Park Cancer Institute (<http://bacpac.med.buffalo.edu>)

NEIGHBORING SEQUENCE INFORMATION:  
 The clone sequenced to the left is AC068592; the clone sequenced to the right is CTD-2096A9. Actual start of this clone is at base position 1 of RP11-564016.

The sequence from base position 103517 to 103554 and from base position 103940 to 104887 is derived from PCR of BAC DNA.

The fidelity of the sequence from base position 48371 to 48772 cannot be guaranteed due to an unresolved imperfect dinucleotide repeat.

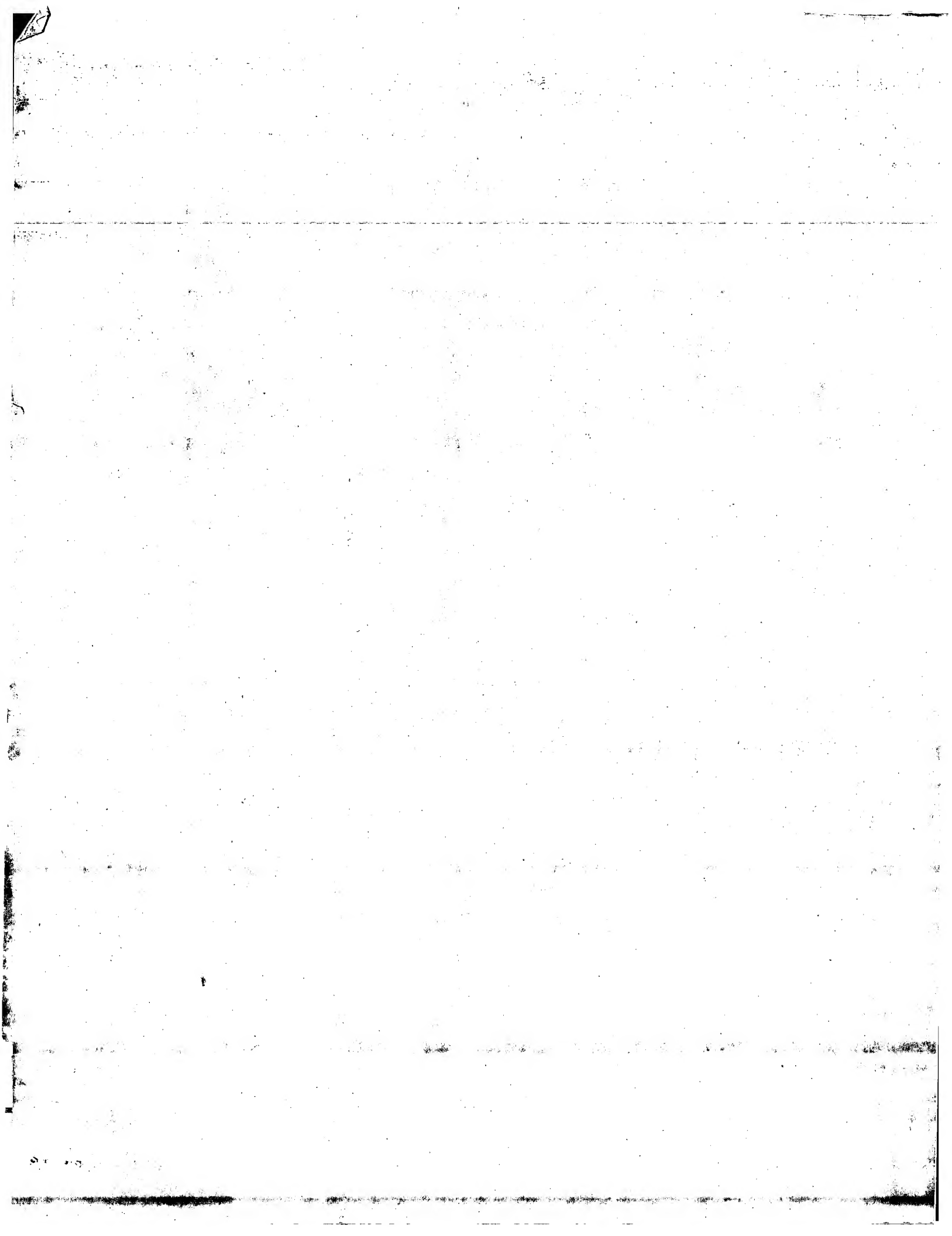
## FEATURES

### source

Location/Qualifiers

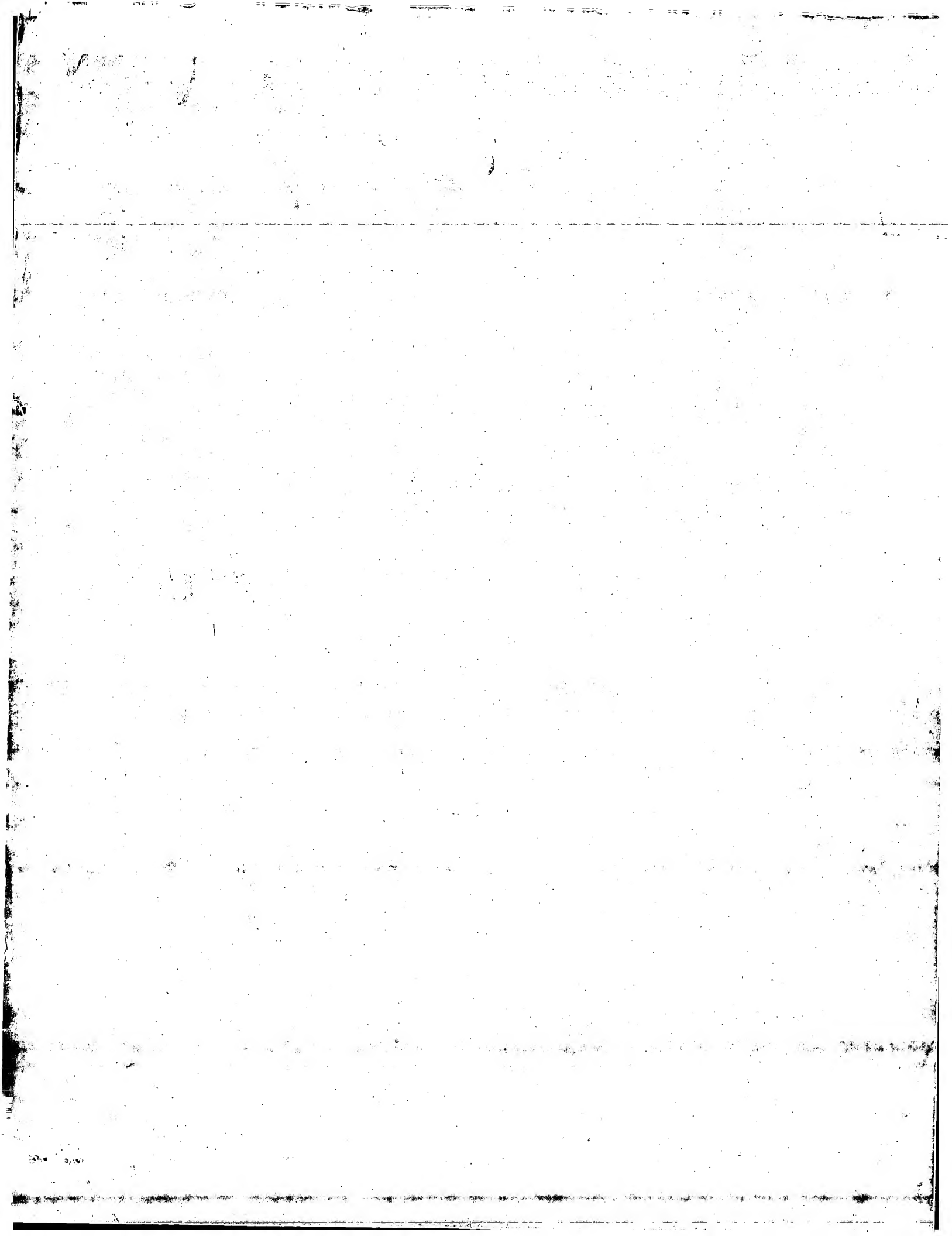
1..128466  
 /organism="Homo sapiens"  
 /db\_xref="taxon:9606"  
 /chromosome="2"  
 /map="2"  
 /clone="RP11-564016"  
 /clone\_id="RPCI-11"  
 291..452  
 /rpt\_family="Alu"  
 1484..1617  
 /rpt\_family="L2"  
 1669..1826  
 /rpt\_family="Alu"  
 1864..2412  
 /rpt\_family="MIR"  
 2413..2591  
 /rpt\_family="L2"  
 2689..2853  
 /rpt\_family="L2"  
 3055..3082  
 /rpt\_family="L2"  
 3085..3392  
 /rpt\_family="L2"  
 3350..3643  
 /rpt\_family="Alu"  
 3644..3717  
 /rpt\_family="CR1"

repeat\_region  
 3718..4183  
 /rpt\_family="Mariner"  
 repeat\_region  
 6123..6293  
 /rpt\_family="CR1"  
 repeat\_region  
 6845..6881  
 /rpt\_family="MIR"  
 repeat\_region  
 6911..6963  
 /rpt\_family="AT\_rich"  
 repeat\_region  
 7910..8008  
 /rpt\_family="AT\_rich"  
 repeat\_region  
 8009..8299  
 /rpt\_family="MIR"  
 repeat\_region  
 8300..8326  
 /rpt\_family="Alu"  
 repeat\_region  
 8327..8599  
 /rpt\_family="MIR"  
 repeat\_region  
 8600..8668  
 /rpt\_family="Alu"  
 repeat\_region  
 9463..9553  
 /rpt\_family="MIR"  
 repeat\_region  
 9890..9923  
 /rpt\_family="TTCA)n"  
 repeat\_region  
 10066..10473  
 /rpt\_family="AT\_rich"  
 repeat\_region  
 10474..10520  
 /rpt\_family="L2"  
 repeat\_region  
 10521..10746  
 /rpt\_family="CA)n"  
 repeat\_region  
 12200..12385  
 /rpt\_family="L2"  
 repeat\_region  
 12452..12489  
 /rpt\_family="MIR"  
 repeat\_region  
 13410..13773  
 /rpt\_family="CACG)n"  
 repeat\_region  
 14076..14317  
 /rpt\_family="MIR"  
 repeat\_region  
 15470..15592  
 /rpt\_family="MIR"  
 repeat\_region  
 15599..15967  
 /rpt\_family="MIR"  
 repeat\_region  
 16227..16617  
 /rpt\_family="MIR"  
 misc\_feature  
 /note="similar to EST AA913068 (NID:93052460) o134e06.s1"  
 repeat\_region  
 16517..16693  
 /rpt\_family="L1"  
 repeat\_region  
 16717..16788  
 /rpt\_family="L1"  
 repeat\_region  
 16988..17009  
 /rpt\_family="AT\_rich"  
 repeat\_region  
 18369..18509  
 /rpt\_family="MER103"  
 repeat\_region  
 18594..18644  
 /rpt\_family="TG)n"  
 repeat\_region  
 19032..19193  
 /rpt\_family="MER1\_type"  
 repeat\_region  
 19359..19668  
 /rpt\_family="L2"  
 repeat\_region  
 22454..22498  
 /rpt\_family="GA-rich"  
 repeat\_region  
 25334..25644  
 /rpt\_family="Alu"  
 repeat\_region  
 25647..25768  
 /rpt\_family="MIR"  
 repeat\_region  
 25770..25888  
 /rpt\_family="L2"  
 repeat\_region  
 25893..26193  
 /rpt\_family="Alu"  
 misc\_feature  
 /note="similar to EST AA578847 (NID:92357031) nk65a09.s1"  
 repeat\_region  
 26368..26442  
 /rpt\_family="MIR"  
 repeat\_region  
 26502..26568  
 /rpt\_family="MIR"

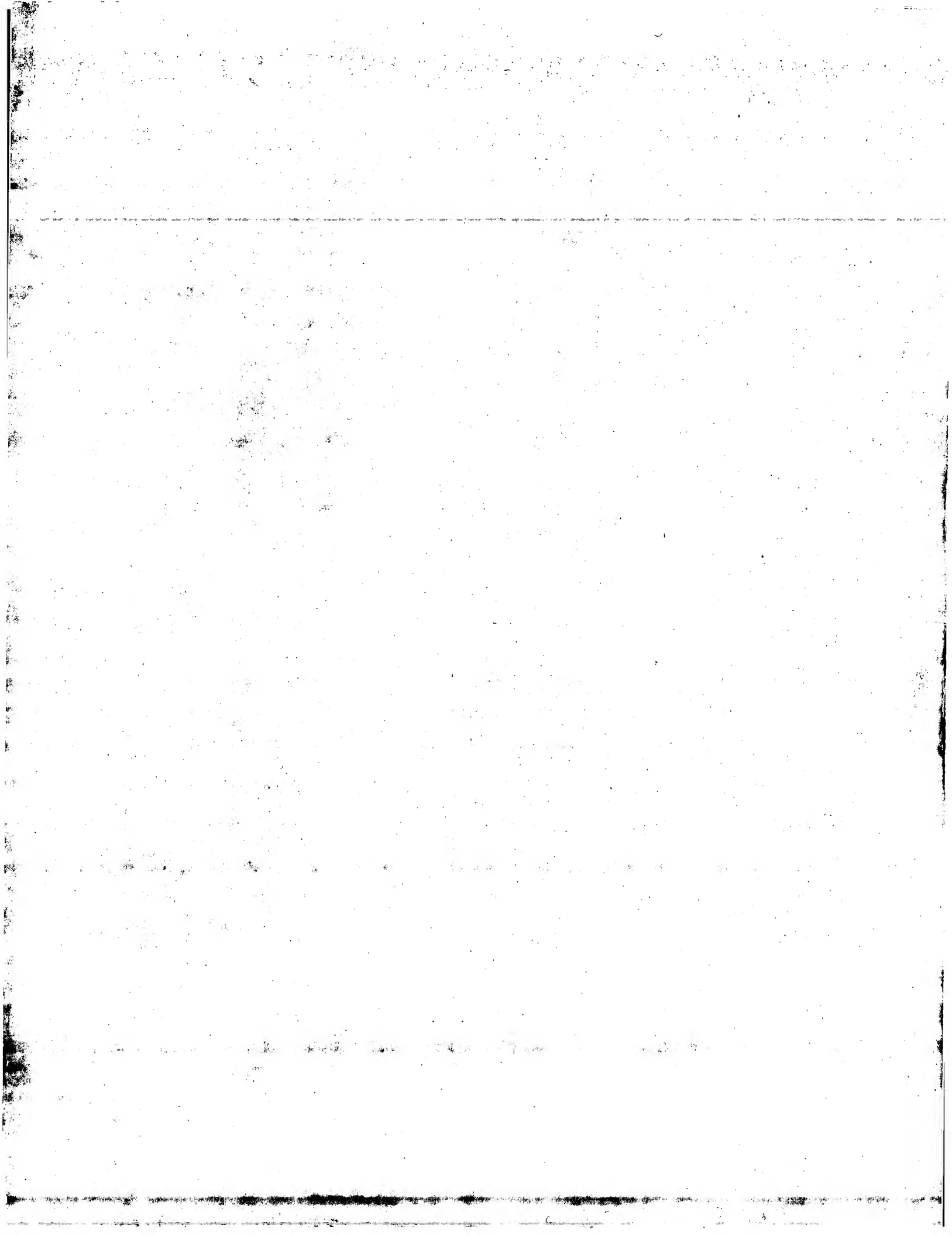


Query Match	93.7%	Score 1309;	DB 9;	length 128466;
. Best Local Similarity	99.6%;	Pred. No. 0;		
Matches 1312;	Conservative	0;	Mismatches 5;	Indels 0;
			Gaps	0;

Oy	779		AAACCTGTGGGCGTGTAGCAAGCTCAGAGCTATGTCCTCCACTGGGAGCGTGCCCTCC	838.
Dd	114515	CACCCTGTGGGCGTGTAGCAAGCTCAGAGCTATGTCCTCCACTGGGAGCGTGCCCTCC		114574
Oy	839	CTGGGACGGTTGCCCTGGGCAGCCCCCATCTACTGTGTTCATTAGTGTAGATAATAGTAATAA		898
Dd	114575	CTGGGACGGTTGCCCTGGGCAGCCCCCATCTACTGTGTTCATTAGTGTAGATAATAGTAATAA		114634
Oy	899	GCCCCCTGTGCTGCTGCTGCACAATGCCACAGAGCGGCTGGGAGCGTGCCTGGGGAACATC		958
Dd	114635	GCCCCCTGTGCTGCTGCTGCACAATGCCACAGAGCGGCTGGGAGCGTGCCTGGGGAACATC		114694
Oy	959	CATCGTGAAGTGTCTCTCACGTTTAGTGTGGAACAGAAGAATTGSCGGGGAATGTCTCCAG		1018
Dd	114695	CATCGTGAAGTGTCTCTCACGTTTAGTGTGGAACAGAAGAATTGSCGGGGAATGTCTCCAG		114754
Oy	1019	GATGTGGGTGATTTCTGTACCTGGGAGGCTATCTTGACCTTCCGACAGGGGACACTCTCC		1078
Dd	114755	GATGTGGGTGATTTCTGTACCTGGGAGGCTATCTTGACCTTCCGACAGGGGACACTCTCC		114814
Oy	1079	AGGCGACGCCAGGGGCTCAGGGGGCAGAGTGCACACTTACGCTTAGGCCCAAAGCTGGGGTC		1138
Dd	114815	AGGCGACGCCAGGGGCTCAGGGGGCAGAGTGCACACTTACGCTTAGGCCCAAAGCTGGGGTC		114874
Oy	1139	AGGAGCAGAGTGTGGTTGAGCCACAGAACCTGGGGCGGGGAGTGGGAGCGGCTTCTGAC		1198
Dd	114875	AGGAGCAGAGTGTGGTTGAGCCACAGAACCTGGGGCGGGGAGTGGGAGCGGCTTCTGAC		114934
Oy	1199	CTCATTTGCTTTCATGAAAGCCCTCAAAAGCAGCCAAAACAGAGCTTTCGCCCTCTGGA		1258
Dd	114935	CTCATTTGCTTTCATGAAAGCCCTCAAAAGCAGCCAAAACAGAGCTTTCGCCCTCTGGA		114994
Oy	1259	GTTTGAAATATCCAGAAATTTTTTGATCTCTGTGTGGTTAAATTTTATTTTGTAAAAA		1318
Dd	114995	GTTTGAAATATCCAGAAATTTTTTGATCTCTGTGTGGTTAAATTTTATTTTGTAAAAA		115054
Oy	1319	ATAAAATTTAAATTAGTTAATAAAATATGATGTTTCAAGCAAACTCTCCCTAAAAAAA		1375
Dd	115055	ATAAAATTTAAATTAGTTAATAAAATATGATGTTTCAAGCAAACTCTCCCTAAAAAAA		115111
RESULT 3				
LOCUS	BC007901	1493 bp	mRNA	linear PRI 12-JUL-2001
DEFINITION	Homo sapiens, clone IMAGE:4139786, mRNA, partial cds.			
ACCESSION	BC007901			
VERSION	BC007901.1 GI:14043927			
KEYWORDS				
SOURCE	Homo sapiens.			
ORGANISM	Homo sapiens.			
	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;			
	Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.			
	1 (bases 1 to 1493)			
REFERENCE	Strausberg, R.			
AUTHORS	Direct Submission			
TITLE	Submitted (11-MAY-2001) National Institutes of Health, Mammalian			
JOURNAL	Gene Collection (MGC), Cancer Genomics Office, National Cancer			
	Institute, 31 Center Drive, Room 11A03, Bethesda, MD 20892-2550,			
	USA			
	NIH-MGC Project URL: http://mgc.nci.nih.gov			
REMARK	Contact: MGC help desk			
COMMENT	Email: cga@nci.nih.gov			
	Tissue Procurement: ATCC			
	cDNA Library Preparation: Rubin Laboratory			
	cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LMNL)			
	DNA Sequencing by: National Institutes of Health Intramural			
	Sequencing Center (NISC),			
	Galthersburg, Maryland:			
	Web site: http://www.nisc.nih.gov/			
	Contact: nisc.mgc@nih.gov			
	nisc.mgc@nih.gov			
	Benjamin, B., Blakesley, R.W., Bouffard, G.G., Brinkley, C., Brooks, S.,			







SO Sequence 9839 BP; 2292 A; 2702 C; 2654 G; 2191 T; 0 other;

Matches 1308; Conservative 0; Mismatches 2; Indels 2; Gaps 2;

Db	9249	CAACCTGTGGGCTGTGTAGCAAGCTCAGAGCTAGTCTCCCACT-GGGGCTGTGCCCTCC	9307
Qy	839	CTGGGACCGTTCCGTGGGACAGCCCATCATCTGTGTTCAATAGTGTAGAAATGTAGCTAA	898
Db	9308	CTGGGACCGTTCCGTGGGACAGCCCATCATCTGTGTTCAATAGTGTAGAAATGTAGCTAA	9367
Qy	899	GCCCCGTGCGTGGCGTGCATGAGCCACACAGACCGGTGGGGGTGCTGTGGGACAATC	958
Db	9368	GCCCCGTGCGTGGCGTGCATGAGCCACACAGACCGGTGGGGGTGCTGTGGGACAATC	9427
Qy	959	CATCGTGAAGTGTCTCTCAGCTTAGTAGGTGAGGACAGGACCTTGGCCGGGAGATGCTCCAG	1018
Db	9428	CATCGTGAAGTGTCTCTCAGCTTAGTAGGTGAGGACAGGACCTTGGCCGGGAGATGCTCCAG	9487
Qy	1019	GATGTGGGTGATTCGTGTAACCTGGGAGAGCTTCTCTGACCTCCGACAGGGGACACTCC	1078
Db	9488	GATGTGGGTGATTCGTGTAACCTGGGAGAGCTTCTCTGACCTCCGACAGGGGACACTCC	9547
Qy	1079	AGGCCAGCCAGGGGTCAGGGGACAGAGTGCACACTCAGCATGAGCCAAAGCTGGGTC	1138
Db	9548	AGGCCAGCCAGGGGTCAGGGGACAGAGTGCACACTCAGCATGAGCCAAAGCTGGGTC	9607
Qy	1139	AGGAGCAGGTGTGTGCTTGAGCCAGACCTTGGGGCGGGGTGGGGCCCTTTCTGC	1198
Db	9608	AGGAGCAGGTGTGTGCTTGAGCCAGACCTTGGGGCGGGGTGGGGCCCTTTCTGC	9667
Qy	1199	CTCATTTGCTTTCAATGAAGACCTCAAGACGCCAAACAGAGCTTCCCTTCCTGCA	1258
Db	9668	CTCATTTGCTTTCAATGAAGACCTCAAGACGCCAAACAGAGCTTCCCTTCCTGCA	9727
Qy	1259	GTTTGAATATCAGAACTTTTGTGACTCTGTGTGGTAAATGTATTTTGTATAAA	1318
Db	9728	GTTTGAATATCAGAACTTTTGTGACTCTGTGTGGTAAATGTATTTTGTATAAA	9787
Qy	1319	ATAAATTAATATTAGTAAATAAATGATGTTTGCACAGCAACTCTTCCCTTA	1370
Db	9788	ATAAATTAATATTAGTAAATAAATGATGTTTGCACAGCAACTCTTCCCTTA	9839

AAI63402 standard; DNA; 9839 BP

22-OCT-2001 (first entry)

KM secondary, therapy, anticancer, immunosuppressive; antitubercular; antihemorrhagic  
KM antiproliferative, cytostatic, cardiant; vasotropic; cerebroprotective  
KM neuroprotective; antibacterial; vitinoid; fungicide;  
KM ophthalmologic; antiallergic; hepatotropic; antidiabetic;  
KM antinflammatory; antitumor; vulnerary; anticonvulsant; antiparasitic;  
KM gene therapy; cancer; immune disorder; cardiovascular disorder;  
KM neurological; disease; infection; ds.

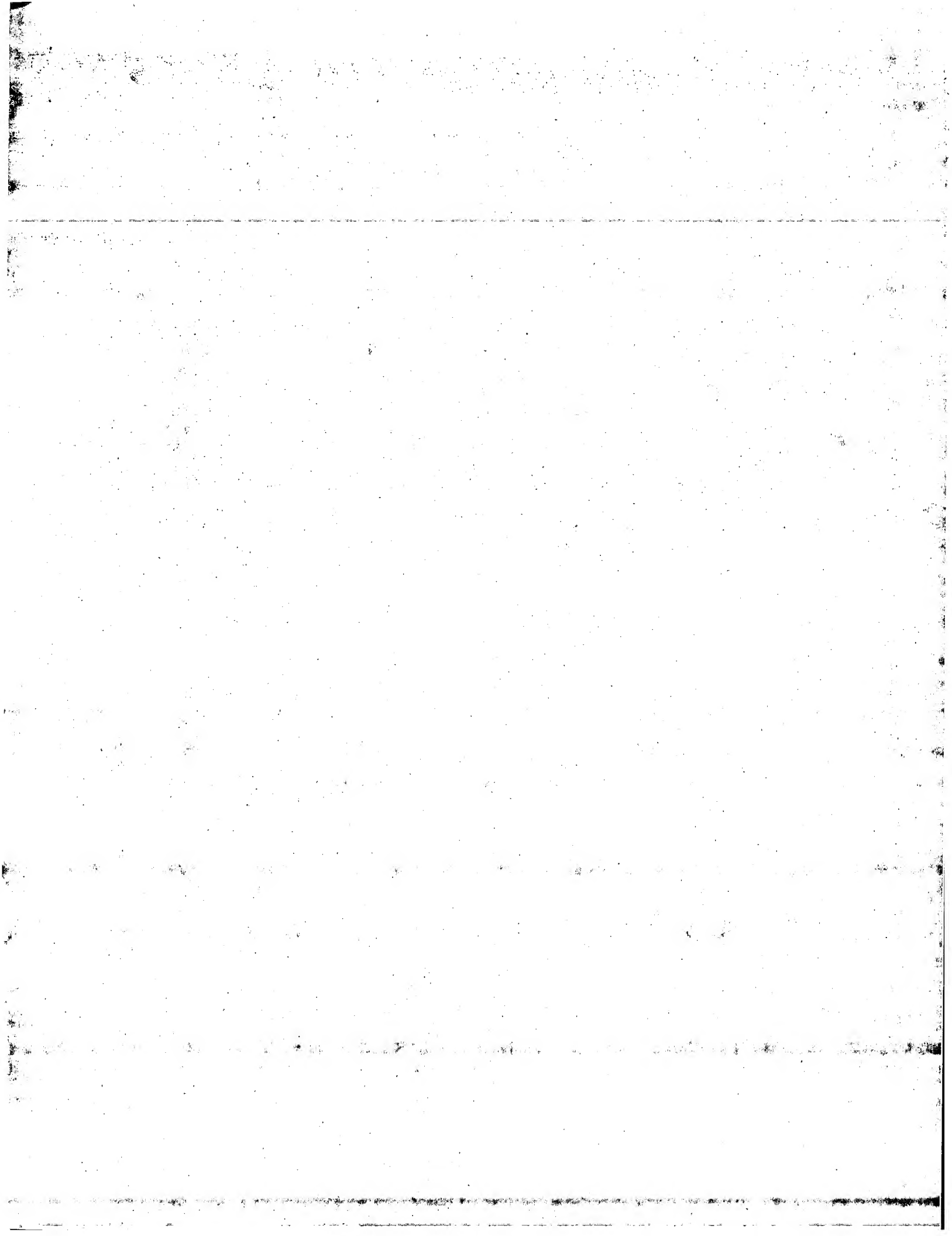
PN WO200155323-A2

17-JAN-2001; 2001WO-US01343

PR 31-JAN-2000; 2000US-0179065

PR 24-FEB-2000; 2000US-0184664  
03 MAR 2000 010000Z  
PR 03-MAR-2000; 2000US-0184664[illegible]

PR : 18-APR-2000; 2000US-0198123





PR 19-MAY-2000; 2000US-0205515.  
PR 07-JUN-2000; 2000US-0209467.  
PR 28-JUN-2000; 2000US-0214886.  
PR 30-JUN-2000; 2000US-0215135.  
PR 07-JUL-2000; 2000US-0216647.  
PR 07-JUL-2000; 2000US-0216880.  
PR 11-JUL-2000; 2000US-0217487.  
PR 14-JUL-2000; 2000US-0218290.  
PR 26-JUL-2000; 2000US-0220963.  
PR 14-AUG-2000; 2000US-0220964.  
PR 14-AUG-2000; 2000US-0224518.  
PR 14-AUG-2000; 2000US-0225213.  
PR 14-AUG-2000; 2000US-0225214.  
PR 14-AUG-2000; 2000US-0225266.  
PR 14-AUG-2000; 2000US-0225286.  
PR 14-AUG-2000; 2000US-0225287.  
PR 14-AUG-2000; 2000US-0225270.  
PR 14-AUG-2000; 2000US-0225447.  
PR 14-AUG-2000; 2000US-0225757.  
PR 14-AUG-2000; 2000US-0225758.  
PR 18-AUG-2000; 2000US-0225755.  
PR 22-AUG-2000; 2000US-0226279.  
PR 22-AUG-2000; 2000US-0226681.  
PR 22-AUG-2000; 2000US-0226686.  
PR 23-AUG-2000; 2000US-0227182.  
PR 30-AUG-2000; 2000US-0227009.  
PR 01-SEP-2000; 2000US-0228982.  
PR 01-SEP-2000; 2000US-0229287.  
PR 01-SEP-2000; 2000US-0229343.  
PR 01-SEP-2000; 2000US-0229344.  
PR 05-SEP-2000; 2000US-0229345.  
PR 05-SEP-2000; 2000US-0229509.  
PR 06-SEP-2000; 2000US-0229553.  
PR 06-SEP-2000; 2000US-0230437.  
PR 08-SEP-2000; 2000US-0230438.  
PR 08-SEP-2000; 2000US-0231242.  
PR 08-SEP-2000; 2000US-0231243.  
PR 08-SEP-2000; 2000US-0231244.  
PR 08-SEP-2000; 2000US-0231413.  
PR 08-SEP-2000; 2000US-0231414.  
PR 08-SEP-2000; 2000US-0232080.  
PR 08-SEP-2000; 2000US-0232081.  
PR 12-SEP-2000; 2000US-0231968.  
PR 14-SEP-2000; 2000US-0232397.  
PR 14-SEP-2000; 2000US-0232398.  
PR 14-SEP-2000; 2000US-0232399.  
PR 14-SEP-2000; 2000US-0232399.  
PR 14-SEP-2000; 2000US-0232400.  
PR 14-SEP-2000; 2000US-0232401.  
PR 14-SEP-2000; 2000US-0233063.  
PR 14-SEP-2000; 2000US-0233064.  
PR 14-SEP-2000; 2000US-0233065.  
PR 21-SEP-2000; 2000US-0234223.  
PR 21-SEP-2000; 2000US-0234274.  
PR 25-SEP-2000; 2000US-0234997.  
PR 25-SEP-2000; 2000US-0234998.  
PR 26-SEP-2000; 2000US-0235484.  
PR 27-SEP-2000; 2000US-0235834.  
PR 27-SEP-2000; 2000US-0235836.  
PR 29-SEP-2000; 2000US-0236367.  
PR 29-SEP-2000; 2000US-0236367.  
PR 29-SEP-2000; 2000US-0236367.  
PR 29-SEP-2000; 2000US-0236369.  
PR 29-SEP-2000; 2000US-0236370.  
PR 02-OCT-2000; 2000US-0236802.  
PR 02-OCT-2000; 2000US-0237037.  
PR 02-OCT-2000; 2000US-0237038.  
PR 02-OCT-2000; 2000US-0237039.  
PR 13-OCT-2000; 2000US-0239935.  
PR 13-OCT-2000; 2000US-0239937.  
PR 30-OCT-2000; 2000US-0240960.

PR 20-OCT-2000; 2000US-0241221.  
PR 20-OCT-2000; 2000US-0241785.  
PR 20-OCT-2000; 2000US-0241786.  
PR 20-OCT-2000; 2000US-0241787.  
PR 20-OCT-2000; 2000US-0241808.  
PR 20-OCT-2000; 2000US-0241809.  
PR 20-OCT-2000; 2000US-0241826.  
PR 01-NOV-2000; 2000US-0244617.  
PR 08-NOV-2000; 2000US-0246474.  
PR 08-NOV-2000; 2000US-0246475.  
PR 08-NOV-2000; 2000US-0246476.  
PR 08-NOV-2000; 2000US-0246477.  
PR 08-NOV-2000; 2000US-0246478.  
PR 08-NOV-2000; 2000US-0246523.  
PR 08-NOV-2000; 2000US-0246524.  
PR 08-NOV-2000; 2000US-0246525.  
PR 08-NOV-2000; 2000US-0246526.  
PR 08-NOV-2000; 2000US-0246527.  
PR 08-NOV-2000; 2000US-0246527.  
PR 08-NOV-2000; 2000US-0246528.  
PR 08-NOV-2000; 2000US-0246532.  
PR 08-NOV-2000; 2000US-0246609.  
PR 08-NOV-2000; 2000US-0246610.  
PR 08-NOV-2000; 2000US-0246611.  
PR 17-NOV-2000; 2000US-0246613.  
PR 17-NOV-2000; 2000US-0249207.  
PR 17-NOV-2000; 2000US-0249208.  
PR 17-NOV-2000; 2000US-0249209.  
PR 17-NOV-2000; 2000US-0249210.  
PR 17-NOV-2000; 2000US-0249211.  
PR 17-NOV-2000; 2000US-0249212.  
PR 17-NOV-2000; 2000US-0249213.  
PR 17-NOV-2000; 2000US-0249214.  
PR 17-NOV-2000; 2000US-0249215.  
PR 17-NOV-2000; 2000US-0249216.  
PR 17-NOV-2000; 2000US-0249217.  
PR 17-NOV-2000; 2000US-0249218.  
PR 17-NOV-2000; 2000US-0249244.  
PR 17-NOV-2000; 2000US-0249245.  
PR 17-NOV-2000; 2000US-0249246.  
PR 17-NOV-2000; 2000US-0249265.  
PR 17-NOV-2000; 2000US-0249297.  
PR 17-NOV-2000; 2000US-0249299.  
PR 01-DEC-2000; 2000US-0249300.  
PR 01-DEC-2000; 2000US-0250160.  
PR 05-DEC-2000; 2000US-0250391.  
PR 05-DEC-2000; 2000US-0251030.  
PR 05-DEC-2000; 2000US-0251988.  
PR 06-DEC-2000; 2000US-0256719.  
PR 06-DEC-2000; 2000US-0251479.  
PR 08-DEC-2000; 2000US-0251856.  
PR 08-DEC-2000; 2000US-0251856.  
PR 08-DEC-2000; 2000US-0251869.  
PR 08-DEC-2000; 2000US-0251869.  
PR 08-DEC-2000; 2000US-0251899.  
PR 11-DEC-2000; 2000US-0251990.  
PR 11-DEC-2000; 2000US-0254097.  
PR 05-JAN-2001; 2001US-0259678.

(HUMA-) HUMAN GENOME SCI INC.

Rosen CA, Barash SC, Ruben SM;

WPI; 2001-488784/53.

New isolated nucleic acids and polypeptides, useful for diagnosing, creating and/or preventing human diseases and disorders.

Disclosure; SEQ ID NO 717; 564bp + Sequence Listing; English.

The invention relates to novel kidney related polynucleotides (AA162971-AA163793) and the encoded polypeptides (AA162971-AA163793) collectively known as kidney antigens and the use of such kidney antigens for detecting disorders of the kidney, especially kidney cancer and kidney cancer metastases. The polynucleotides and proteins are also



useful for preventing, treating or ameliorating medical conditions  
e.g. by protein or gene therapy. The genes are isolated from a range  
of human tissues disclosed in the specification. The nucleic acids,  
proteins, antibodies and (ant)agonists are useful in the diagnosis,  
treatment and prevention of: (a) cancer, e.g. breast and ovarian cancer,  
and other cancers of the adrenal gland, bone, bone marrow, breast,  
gastrointestinal tract, liver, lung, or urogenital; (b) immune disorders  
e.g. Addison's disease, allergies, autoimmune haemolytic anaemia,  
autoimmune thyroiditis, diabetes mellitus, Crohn's disease, multiple  
sclerosis, rheumatoid arthritis and ulcerative colitis;  
(c) cardiovascular disorders such as myocardial ischaemia; (d) wound  
healing; (e) neurological diseases e.g. cerebral anoxia and epilepsy;  
and (f) infectious diseases such as viral, bacterial, fungal and  
parasitic infections.  
Note: The sequence data for this patent did not form part of the  
printed specification, but was obtained in electronic format directly  
from WIPO at ftp.wipo.int/pub/published\_pct\_sequences.

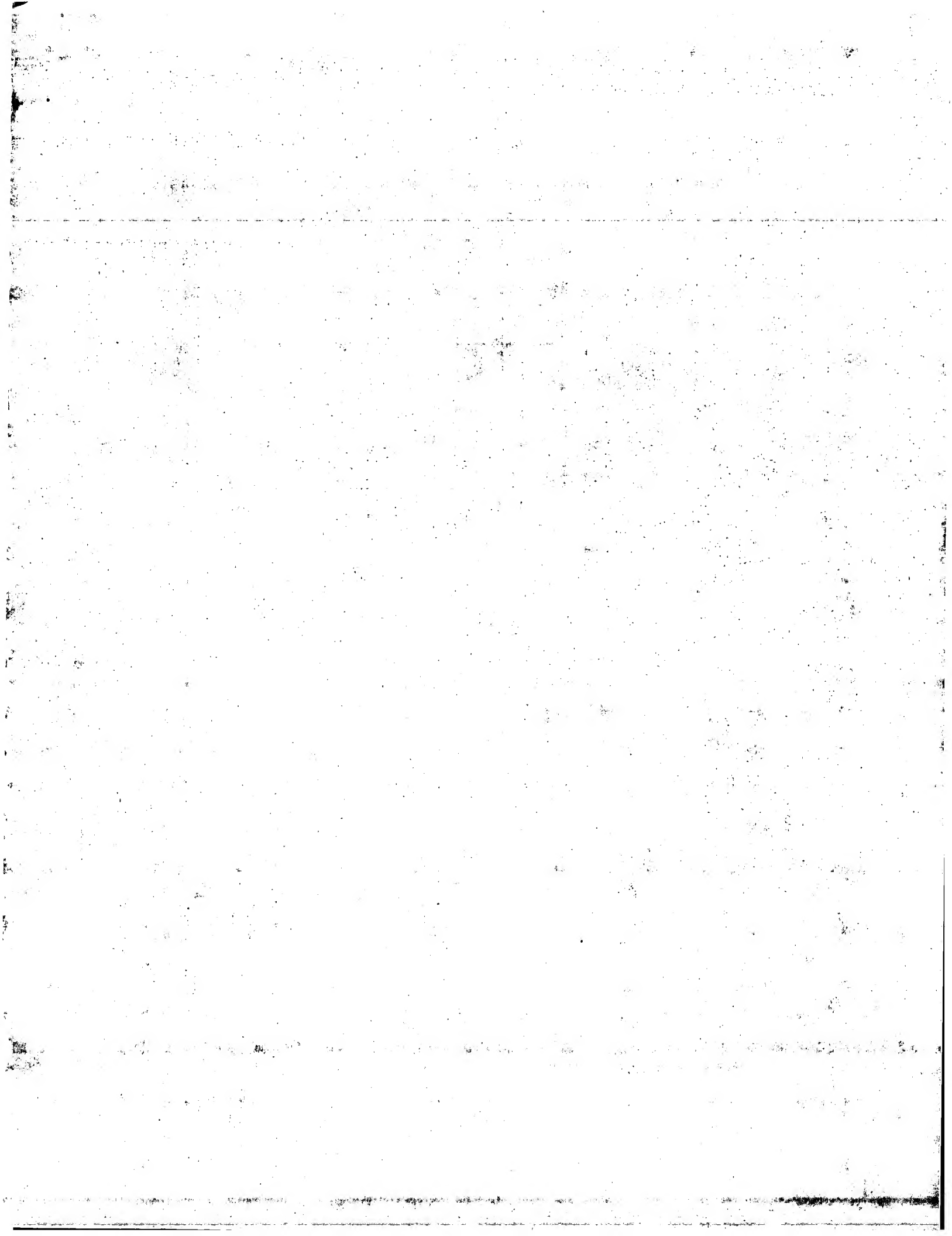
Sequence 9839 BP; 2292 A; 2702 C; 2654 G; 2191 T; 0 other;

Query Match 92.0% Score 1284.8; DB 22; Length 9839;  
Best Local Similarity 99.7%; Pred. No. 1.1e-289;  
Matches 1308; Conservative 0; Mismatches 2; Indels 2; Gaps 2;

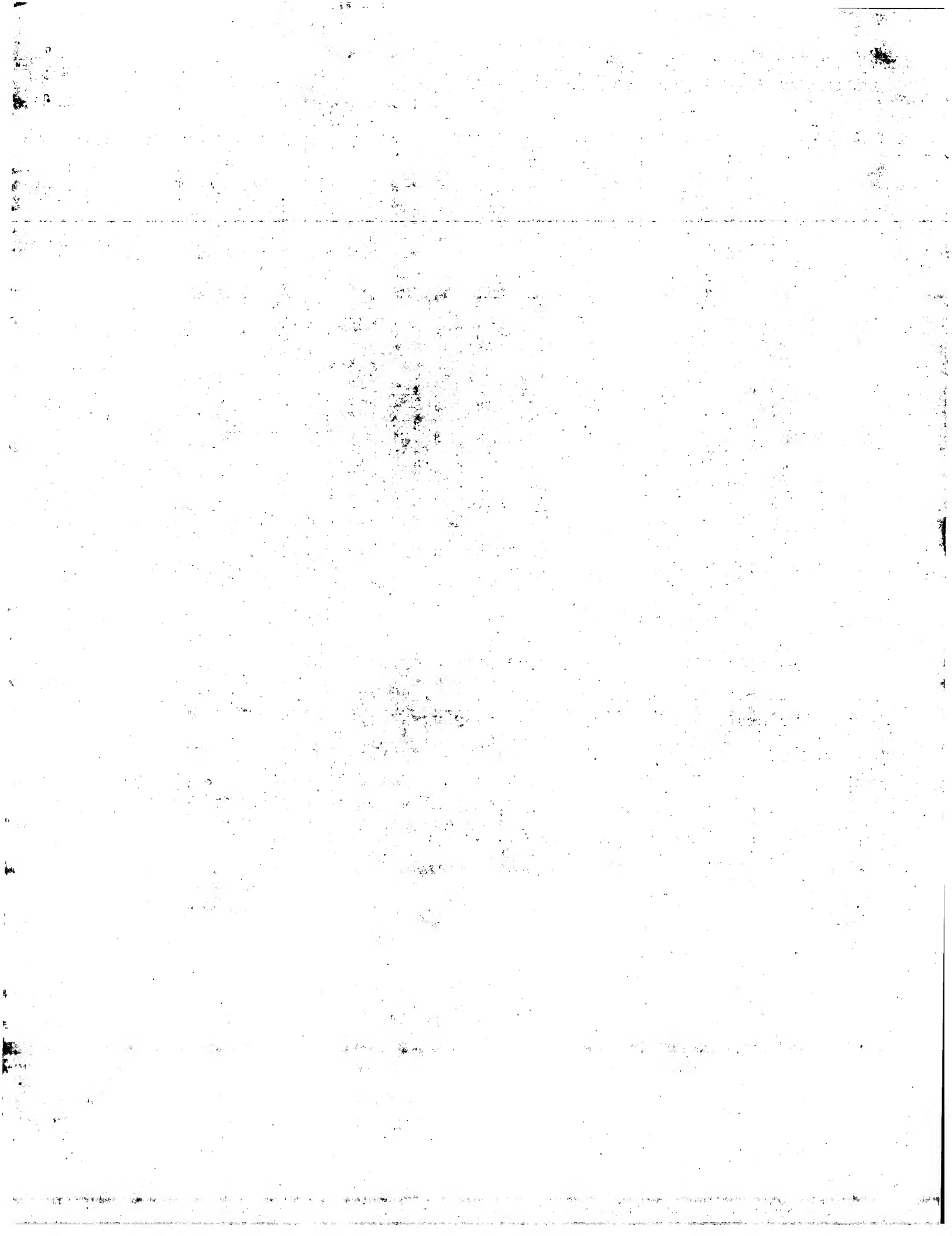
59 ACCGAGTACAGTGTATCCAGACAGACATCCCGGAGAACTACCGCTGTCGCA 118  
8530 ACAGAGTACAGTGTATCCAGACAGACATCCCGGAGAACTACCGCTGTCGCA 8589  
119 TCCATACACAGCGGAGCTGCTCTTCACTGTTCACTGCTGAGGCTGTGATGTC 178  
8590 TCCATACACAGCGGAGCTGCTCTTCACTGTTCACTGCTGAGGCTGTGATGTC 8649  
179 TGTAGAGCAATGCGCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 238  
8650 TGTAGAGCAATGCGCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 8709  
239 GAGCAGTGGAGAGAGCCCTTCAAGGAGATGAGCAGGACCTCTGAGGTTGATGAT 298  
8710 GAGCAGTGGAGAGAGCCCTTCAAGGAGATGAGCAGGACCTCTGAGGTTGATGAT 8769  
299 AGTATATCCCATGCGAGTCAAGGAGGCTGCTGAGTGTAGAGAGAGGTTATGCTG 358  
8770 AGTATATCCCATGCGAGTCAAGGAGGCTGCTGAGTGTAGAGAGAGGTTATGCTG 8829  
359 CTTCAAGGAGTCAAAATTAAGGAGAAATGCTGCTGCTGCTGCTGCTGCTGCTGCT 418  
8830 CTTCAAGGAGTCAAAATTAAGGAGAAATGCTGCTGCTGCTGCTGCTGCTGCTGCT 8889  
419 TGTATCTCTCACTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 478  
8890 TGTATCTCTCACTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 8949  
479 CCAAGTGTCTCTGATCCCAAGAACACATATGTAAGGCTGCTGCTGCTGCTGCTGCT 538  
8950 CCAAGTGTCTCTGATCCCAAGAACACATATGTAAGGCTGCTGCTGCTGCTGCTGCT 9008  
539 AGGCTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 598  
9009 AGGCTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 9068  
599 TTGCACTGCGACAGCATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 658  
9069 TTGCACTGCGACAGCATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 9128  
659 CAGACAGATGTGACAGAGCAAACTGCAATATGCAATATGTTAAATGTAAGTTTAC 718  
9129 CAGACAGATGTGACAGAGCAAACTGCAATATGCAATATGTTAAATGTAAGTTTAC 9188  
719 AGCCTAGCTATGGAATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 778  
9189 AGCCTAGCTATGGAATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 9248  
779 CAACCTGTGGCTGTATAGCAAGAGCTGAGCTAGTCTCCCACTGAGGAGCTGCTGCT 838

9249 CAACCTGTGGCTGTATAGCAAGAGCTGAGGCTAGTCTCCCACT-GGGGCTGCTGCTCC 9307  
839 CTGGAGCGGTTCGCTGGGCGAGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 898  
9308 CTGGAGCGGTTCGCTGGGCGAGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 9367  
899 GCCCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 958  
9368 GCCCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 9427  
959 CATGCTGAGTGTCTCTCACTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1018  
9428 CATGCTGAGTGTCTCTCACTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 9487  
1019 GATGCTGAGTGTCTCTCACTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1078  
9488 GATGCTGAGTGTCTCTCACTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 9547  
1079 AGGCGACCGGAGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1138  
9548 AGGCGACCGGAGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 9607  
1139 AGGCGACCGGAGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1198  
9608 AGGCGACCGGAGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 9667  
1199 CTGATGCTGCTTCAATGAAGGCTCAAGAGCGCAAAACAGGCTTCCCTCTCGA 1258  
9668 CTGATGCTGCTTCAATGAAGGCTCAAGAGCGCAAAACAGGCTTCCCTCTCGA 9727  
1259 GTTGAATATCAGAAATCTTGTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1318  
9728 GTTGAATATCAGAAATCTTGTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 9787  
1319 ATAAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 1370  
9788 ATAAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 9839

RESULT 3  
AAH98172  
ID AAH98172 standard; cDNA, 1846 BP.  
XX  
AC AAH98172;  
XX  
DT 12-OCT-2001 (first entry)  
XX  
DE Human EST-derived coding sequence SEQ ID NO: 29.  
XX  
KW Human; sheep; pig; cow; fruit fly; yeast; hamster; macaque; horse;  
XX tomato; monkey; dog; sea urchin; expressed sequence tag; EST;  
XX diagnostics; forensic test; gene mapping; genetic disorder;  
XX biodiversity; gene therapy; nutrition; ss.  
XX  
OS Homo sapiens.  
XX  
PN WO200154477-A2.  
XX  
PD 02-AUG-2001.  
XX  
PF 25-JAN-2001; 2001WO-US02687.  
XX  
PR 25-JAN-2000; 2000US-0491404.  
XX 17-JUL-2000; 2000US-0617746.  
XX 03-AUG-2000; 2000US-0631451.  
XX 15-SEP-2000; 2000US-0663870.  
XX  
PA (HYSB-) HYSBQ INC.  
XX  
PI Tang YF, Liu C, Zhou P, Qian XB, Wang Z, Chen R, Aundt V,  
XX Cao Y, Dermanac RA, Zhang J, Wehrman T;  
XX



PR	04-FEB-2000	2000US-0186628
PR	24-FEB-2000	2000US-0184664
PR	02-MAR-2000	2000US-0186350
PR	15-MAR-2000	2000US-0189874
PR	17-MAR-2000	2000US-0190016
PR	18-APR-2000	2000US-0198123
PR	19-MAY-2000	2000US-0205515
PR	07-JUN-2000	2000US-0209467
PR	28-JUN-2000	2000US-0214886
PR	30-JUN-2000	2000US-0215153
ER	07-JUL-2000	2000US-0216647
PR	07-JUL-2000	2000US-0216880
PR	11-JUL-2000	2000US-0217487
PR	11-JUL-2000	2000US-0217496
PR	14-JUL-2000	2000US-0218920
PR	26-JUL-2000	2000US-0220963
PR	26-JUL-2000	2000US-0220964
PR	14-AUG-2000	2000US-0224518
PR	14-AUG-2000	2000US-0224519
PR	14-AUG-2000	2000US-0225521
PR	14-AUG-2000	2000US-0225521.4
PR	14-AUG-2000	2000US-0225266
PR	14-AUG-2000	2000US-0225267
PR	14-AUG-2000	2000US-0225268
PR	14-AUG-2000	2000US-0225270
PR	14-AUG-2000	2000US-0225447
PR	14-AUG-2000	2000US-0225577
PR	14-AUG-2000	2000US-0225758
PR	18-AUG-2000	2000US-0226759
PR	18-AUG-2000	2000US-0226759
PR	22-AUG-2000	2000US-0226681
PR	22-AUG-2000	2000US-0226681
PR	22-AUG-2000	2000US-0226686
PR	23-AUG-2000	2000US-0227182
PR	30-AUG-2000	2000US-0228700
PR	30-AUG-2000	2000US-0228924
PR	01-SEP-2000	2000US-0229287
PR	01-SEP-2000	2000US-0229343
PR	01-SEP-2000	2000US-0229344
PR	01-SEP-2000	2000US-0229345
PR	05-SEP-2000	2000US-0229509
PR	05-SEP-2000	2000US-0229513
PR	06-SEP-2000	2000US-0230438
PR	08-SEP-2000	2000US-0231342
PR	08-SEP-2000	2000US-0231243
PR	08-SEP-2000	2000US-0231443
PR	08-SEP-2000	2000US-0231444
PR	08-SEP-2000	2000US-0231414
PR	08-SEP-2000	2000US-0231414
PR	08-SEP-2000	2000US-0231480
PR	08-SEP-2000	2000US-0232081
PR	12-SEP-2000	2000US-0231368
PR	14-SEP-2000	2000US-0232397
PR	14-SEP-2000	2000US-0232398
PR	14-SEP-2000	2000US-0232399
PR	14-SEP-2000	2000US-0232400
PR	14-SEP-2000	2000US-0232401
PR	14-SEP-2000	2000US-0233063
PR	14-SEP-2000	2000US-0233065
PR	21-SEP-2000	2000US-0233423
PR	21-SEP-2000	2000US-0234472
PR	25-SEP-2000	2000US-0234997
PR	25-SEP-2000	2000US-0234998
PR	26-SEP-2000	2000US-0234998
PR	27-SEP-2000	2000US-0235834
PR	27-SEP-2000	2000US-0235836
PR	29-SEP-2000	2000US-0236367
PR	29-SEP-2000	2000US-0236368
PR	29-SEP-2000	2000US-0236369
PR	29-SEP-2000	2000US-0236370
PR	02-OCT-2000	2000US-0237602
PR	02-OCT-2000	2000US-0237603



02-OCT-2000; 2000US-0237038.  
 02-OCT-2000; 2000US-0237039.  
 02-OCT-2000; 2000US-0237040.  
 13-OCT-2000; 2000US-0239335.  
 13-OCT-2000; 2000US-0239337.  
 20-OCT-2000; 2000US-0240360.  
 20-OCT-2000; 2000US-0241221.  
 20-OCT-2000; 2000US-0241785.  
 20-OCT-2000; 2000US-0241786.  
 20-OCT-2000; 2000US-0241787.  
 20-OCT-2000; 2000US-0241808.  
 20-OCT-2000; 2000US-0241809.  
 01-NOV-2000; 2000US-0244826.  
 01-NOV-2000; 2000US-0244827.  
 08-NOV-2000; 2000US-0246474.  
 08-NOV-2000; 2000US-0246475.  
 08-NOV-2000; 2000US-0246476.  
 08-NOV-2000; 2000US-0246477.  
 08-NOV-2000; 2000US-0246478.  
 08-NOV-2000; 2000US-0246523.  
 08-NOV-2000; 2000US-0246524.  
 08-NOV-2000; 2000US-0246525.  
 08-NOV-2000; 2000US-0246526.  
 08-NOV-2000; 2000US-0246527.  
 08-NOV-2000; 2000US-0246528.  
 08-NOV-2000; 2000US-0246532.  
 08-NOV-2000; 2000US-0246609.  
 08-NOV-2000; 2000US-0246610.  
 08-NOV-2000; 2000US-0246611.  
 17-NOV-2000; 2000US-0249207.  
 17-NOV-2000; 2000US-0249208.  
 17-NOV-2000; 2000US-0249209.  
 17-NOV-2000; 2000US-0249210.  
 17-NOV-2000; 2000US-0249211.  
 17-NOV-2000; 2000US-0249212.  
 17-NOV-2000; 2000US-0249213.  
 17-NOV-2000; 2000US-0249214.  
 17-NOV-2000; 2000US-0249215.  
 17-NOV-2000; 2000US-0249216.  
 17-NOV-2000; 2000US-0249217.  
 17-NOV-2000; 2000US-0249218.  
 17-NOV-2000; 2000US-0249244.  
 17-NOV-2000; 2000US-0249245.  
 17-NOV-2000; 2000US-0249264.  
 17-NOV-2000; 2000US-0249265.  
 17-NOV-2000; 2000US-0249297.  
 17-NOV-2000; 2000US-0249299.  
 01-DEC-2000; 2000US-0249300.  
 01-DEC-2000; 2000US-0250160.  
 01-DEC-2000; 2000US-0250391.  
 05-DEC-2000; 2000US-0251030.  
 05-DEC-2000; 2000US-0251988.  
 05-DEC-2000; 2000US-0256719.  
 06-DEC-2000; 2000US-0251479.  
 08-DEC-2000; 2000US-0251856.  
 08-DEC-2000; 2000US-0251868.  
 08-DEC-2000; 2000US-0251869.  
 08-DEC-2000; 2000US-0251989.  
 08-DEC-2000; 2000US-0251990.  
 11-DEC-2000; 2000US-0254097.  
 05-JAN-2001; 2001US-0259678.  
 (HUMA-) HUMAN GENOME SCI INC.  
 Rosen CA, Barash SC, Ruben SM;  
 WPI; 2001-488784/53.  
 New isolated nucleic acids and polypeptides, useful for diagnosing,  
 treating and/or preventing human diseases and disorders -  
 Disclosure; SEQ ID NO 717; 564pp + Sequence Listing; English.

XX The invention relates to novel kidney related polynucleotides  
 CC (AA162971-AA163793) and the encoded polypeptides (AA162417-AA162691)  
 CC collectively known as kidney antigens and the use of such kidney antigens  
 CC for detecting disorders of the kidney, especially kidney cancer and  
 CC kidney cancer metastases. The polynucleotides and proteins are also  
 CC useful for preventing, treating or ameliorating medical conditions  
 CC e.g. by protein or gene therapy. The genes are isolated from a range  
 CC of human tissues disclosed in the specification. The nucleic acids  
 CC proteins, antibodies and (anti)agonists are useful in the diagnosis,  
 CC treatment and prevention of: (a) cancer, e.g. breast and ovarian cancer,  
 CC and other cancers of the adrenal gland, bone, bone marrow, breast,  
 CC gastrointestinal tract, liver, lung, or urogenital; (b) immune disorders  
 CC e.g. Addison's disease, allergies, autoimmune haemolytic anaemia,  
 CC autoimmune thyroiditis, diabetes mellitus, Crohn's disease, multiple  
 CC sclerosis, rheumatoid arthritis and ulcerative colitis;  
 CC (c) cardiovascular disorders such as myocardial ischaemias; (d) wound  
 CC healing; (e) neurological diseases e.g. cerebral anoxia and epilepsy;  
 CC and (f) infectious diseases such as viral, bacterial, fungal and  
 CC parasitic infections.  
 CC Note: The sequence data for this patent did not form part of the  
 CC printed specification, but was obtained in electronic format directly  
 CC from WPI at ftp.wipo.int/pub/published\_pat\_sequences.  
 XX  
 SQ Sequence 9839 BP; 2292 A; 2702 C; 2654 G; 2191 T; 0 other;  
 Query Match 72.7%; Score 1016; DB 22; Length 9839;  
 Best Local Similarity 99.8%; Pred. No. 0;  
 Matches 1106; Conservative 0; Mismatches 1; Indels 2; Gaps 2;  
 QY 62 GAGTACAGTGTATCCAGACAGACATCCCGAGAGATACCGCTGCGGCATCC 121  
 DB 8533 GAGTACAGTGTATCCAGACAGACATCCCGAGAGATACCGCTGCGGCATCC 8592  
 QY 122 TACCAACACGGGAGCTGCTCTTCACTGTTCAACTGCTGAGGCTGTGATCTGT 181  
 DB 8593 TACCAACACGGGAGCTGCTCTTCACTGTTCAACTGCTGAGGCTGTGATCTGT 8652  
 QY 182 GAGAGCATGCGCCAGTGTGCGGCTTTGTGTACCAACAGACCACTGACAGTGG 241  
 DB 8653 GAGAGCATGCGCCAGTGTGCGGCTTTGTGTACCAACAGACCACTGACAGTGG 8712  
 QY 242 CCAATGGGAGAGCCCTTCAAGGAGATGACAGACCTCTGAGGCTGATATAGTCT 301  
 DB 8713 CCAATGGGAGAGCCCTTCAAGGAGATGACAGACCTCTGAGGCTGATATAGTCT 8772  
 QY 302 GATCCCCATGGAAGTCAAGAGGAGGCTGAGGTGATGAGAGAGATATAGTCT 361  
 DB 8773 GATCCCCATGGAAGTCAAGAGGAGGCTGAGGTGATGAGAGAGATATAGTCT 8832  
 QY 362 CAAGGAGTCAAAATTAAGGAGAAATGCTTCCCTCCAGAAAGAGAAATCATCAGCCCTGT 421  
 DB 8833 CAAGGAGTCAAAATTAAGGAGAAATGCTTCCCTCCAGAAAGAGAAATCATCAGCCCTGT 8892  
 QY 422 TACCTCTACCTTGCCTCCCGAGGTGCGAGCTGCTCTTTCAAGATGATGAGAGCA 481  
 DB 8893 TACCTCTACCTTGCCTCCCGAGGTGCGAGCTGCTCTTTCAAGATGATGAGAGCA 8952  
 QY 482 AGTGGTCCCTGATCCCAACAGACCAATATGTAATGCTTGGCTGAGTGAAGTGAAG 541  
 DB 8953 AGTGGTCCCTGATCCCAACAGACCAATATGTAATGCTTGGCTGAGTGAAGTGAAG 9011  
 QY 542 GCTGGCTGACAGCTGATCTATCTCTGACAGCTGCTGCTGCTGCTGAGAGTGAAGT 601  
 DB 9012 GCTGGCTGACAGCTGATCTATCTCTGACAGCTGCTGCTGCTGCTGAGAGTGAAGT 9071  
 QY 602 CACTGGAGACATGATGATCACTGAGAGCCCTGAGAGCAAAAGCTAACTCCAGACAG 661  
 DB 9072 CACTGGAGACATGATGATCACTGAGAGCCCTGAGAGCAAAAGCTAACTCCAGACAG 9131  
 QY 662 ACAGATGACCAAGACAAAGTGCATTAATGCAATATGTTAAATGATGATTAACAGC 721  
 DB 9132 ACAGATGACCAAGACAAAGTGCATTAATGCAATATGTTAAATGATGATTAACAGC 9191





```

OY 722 CTAGCTATGGAGCTGCTGCTCTCTAGTCCAGGAATCATGAGGGATATGACTGCTCTCCAA 781
DB 9192 CTAGCTATGGAGCTGCTGCTCTCTAGTCCAGGAATCATGAGGGATATGACTGCTCTCCAA 9251
OY 782 CCTGTGGAGCTGTAAAGCAAGCTCAGGCTTATGCTCCCACTGGGGGCTGTCGCCCTCCG 841
DB 9252 CCTGTGGAGCTGTAAAGCAAGCTCAGGCTTATGCTCCCACTGGGGGCTGTCGCCCTCCG 9310
OY 842 GGAAGCTTCCGTGGGAGAGCCCATCATCTGTGTTCAATAGTGTAGATTAAGCTTAAAGCC 901
DB 9311 GGAAGCTTCCGTGGGAGAGCCCATCATCTGTGTTCAATAGTGTAGATTAAGCTTAAAGCC 9370
OY 902 CCTGTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 961
DB 9371 CCTGTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 9430
OY 962 CGTGAAGTGTCTCTCAAGCTTAAAGCTTGAAGCAAGCACTTGGGGGGGATGCTCCAGAT 1021
DB 9431 CGTGAAGTGTCTCTCAAGCTTAAAGCTTGAAGCAAGCACTTGGGGGGGATGCTCCAGAT 9490
OY 1022 GTGGGATATCTGTAAGCTGAGGAGGCTATCTGTAAGCTGTAAGCTGTAAGCTGTAAG 1081
DB 9491 GTGGGATATCTGTAAGCTGAGGAGGCTATCTGTAAGCTGTAAGCTGTAAGCTGTAAG 9550
OY 1082 CCAAGCCAGAGGAGTCAAGGAGGAGTCAAGCTGTAAGCTGTAAGCTGTAAGCTGTAAG 1141
DB 9551 CCAAGCCAGAGGAGTCAAGGAGGAGTCAAGCTGTAAGCTGTAAGCTGTAAGCTGTAAG 9610
OY 1142 GAGCAGGTGTGTGTTGAGCAGAGCTTGGGGGCTGTTGGGGGCTTGTGCTGCTC 1201
DB 9611 GAGCAGGTGTGTGTTGAGCAGAGCTTGGGGGCTGTTGGGGGCTTGTGCTGCTC 9670
OY 1202 ATTTGCTTCAATGAAAGCTTCAAGCAAGCAAGCAAGCAAGCAAGCAAGCAAGCAAG 1261
DB 9671 ATTTGCTTCAATGAAAGCTTCAAGCAAGCAAGCAAGCAAGCAAGCAAGCAAGCAAG 9730
OY 1262 TGAATATCCAGAACTTTTGTGTAAGCTTGTGTTAAATTTGTTAAATTTGTTAAAT 1321
DB 9731 TGAATATCCAGAACTTTTGTGTAAGCTTGTGTTAAATTTGTTAAATTTGTTAAAT 9790
OY 1322 AAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTA 1370
DB 9791 AAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTA 9839

```

## RESULT 3

AAH98172 standard; cDNA; 1846 BP.

AAH98172:

12-OCT-2001 (first entry)

Human EST-derived coding sequence SEQ ID NO: 29.

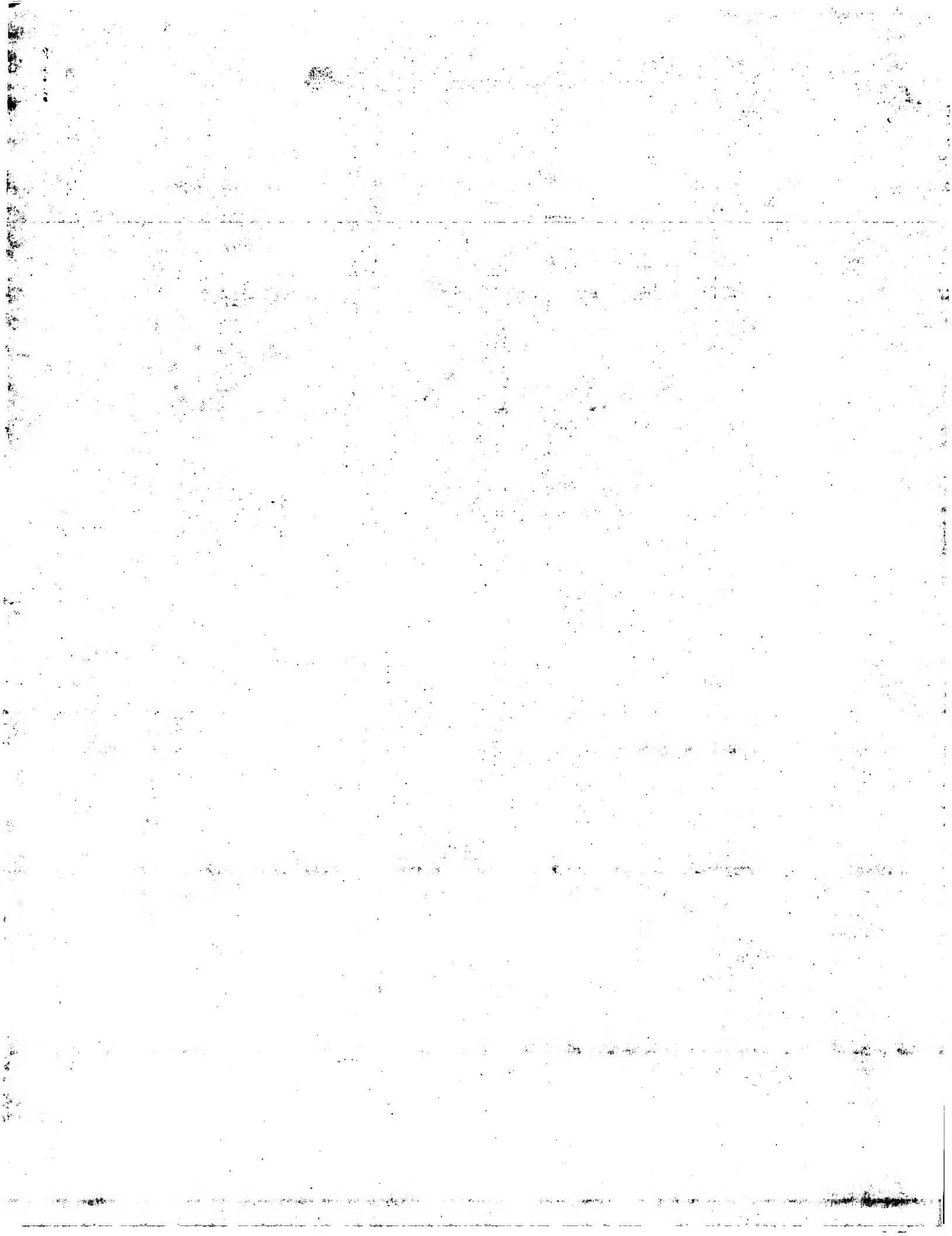
Human, sheep; pig; cow; fruit fly; Yeast; hamster; macaque; horse;  
 tomato; monkey; dog; sea urchin; expressed sequence tag; EST;  
 diagnostics; forensic; gene mapping; genetic disorder;  
 biodiversity; gene therapy; nutrition; ss.

Homo sapiens.  
 WO200154477-A2.  
 02-AUG-2001.  
 25-JAN-2001; 2001WO-US02687.  
 25-JAN-2000; 2000US-0491404.  
 17-JUL-2000; 2000US-0617746.  
 03-AUG-2000; 2000US-0631451.  
 03-SEP-2000; 2000US-0663870.

```

XX (HYSE-) HYSEQ INC.
PA Tang YT, Liu C, Zhou P, Qian XB, Wang Z, Chen R, Aundi V;
XX Cao Y, Dermanac RA, Zhang J, Werhman T;
PI WPI; 2001-476164/51.
XX P-PSDB; AAM23513.
XX DR
XX PT Isolated polypeptide for treatment of diseases, diagnostics, raising
XX antibodies and research use -
PS Claim 1; Page 202-203; 1275pp; English.
XX
XX The present invention provides the protein and coding sequences of novel
XX proteins from a variety of organisms, including human, dog, cat, horse,
XX cow, pig, hamster, monkey, macaque, yeast, bacteria, fruit fly, sea
XX urchin and tomato. These were derived from expressed sequence tags (ESTs)
XX from the organism of interest. They can be used in diagnostics,
XX forensics, gene mapping, identification of mutations, to assess
XX biodiversity and for nutritional purposes. The present sequence is a cDNA
XX of the invention.
SQ Sequence 1846 BP; 397 A; 526 C; 538 G; 385 T; 0 other;
XX
XX Query Match 63.6%; Score 888; DB 22; Length 1846;
XX Best Local Similarity 99.9%; Pred. No. 0;
XX Matches 938; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
OY 442 CAGGTGGCAGCTGCTGCTTTTCAAGCTGAGTGAAGCAAGTGTGCTGATCCCAACA 501
DB 908 CAGGTGGCAGCTGCTGCTTTTCAAGCTGAGTGAAGCAAGTGTGCTGATCCCAACA 967
OY 502 AGACCAATATGTAAGGCTCTGCTGTAAGCTTCTGAGGCTGCTGCTGCTGCTGCTGCT 561
DB 968 AGACCAATATGTAAGGCTCTGCTGTAAGCTTCTGAGGCTGCTGCTGCTGCTGCTGCT 1027
OY 562 ATCTCAGCAGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 621
DB 1028 ATCTCAGCAGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1087
OY 622 ACTTGGGAAGCTTGGCAAGCAAGCTTCAATCCAGACAGACAGATGTACCAAGCAAA 681
DB 1088 ACTTGGGAAGCTTGGCAAGCAAGCTTCAATCCAGACAGACAGATGTACCAAGCAAA 1147
OY 682 CGTGCAATATGCAAGCTTAAATGTAGTTTCAAGCTTCAAGCTTCAAGCTTCAAGCT 741
DB 1148 CGTGCAATATGCAAGCTTAAATGTAGTTTCAAGCTTCAAGCTTCAAGCTTCAAGCT 1207
OY 742 TCTTAGTCCAGGAATCATGAGGGATATGACTGCTTCCAAAGCTTGGGGCTGTAAGCAAG 801
DB 1208 TCTTAGTCCAGGAATCATGAGGGATATGACTGCTTCCAAAGCTTGGGGCTGTAAGCAAG 1267
OY 802 CTCAGGCTATGCTTCCCACTGGGGGCTGAGCCCTTCCCTGAGGACGGTTCCGTGGAGGCC 861
DB 1268 CTCAGGCTATGCTTCCCACTGGGGGCTGAGCCCTTCCCTGAGGACGGTTCCGTGGAGGCC 1327
OY 862 CCATCACTGCTGCTCAATAGTGTAGATTAAGCTTCAAGCTTCAAGCTTCAAGCTTCAAG 921
DB 1328 CCATCACTGCTGCTCAATAGTGTAGATTAAGCTTCAAGCTTCAAGCTTCAAGCTTCAAG 1387
OY 922 TGCACAGCAGGCGGTGGGGGCTGCTGCTGAGGACATTCATCTGAGGATGTTCTCAAGCT 981
DB 1388 TGCACAGCAGGCGGTGGGGGCTGCTGCTGAGGACATTCATCTGAGGATGTTCTCAAGCT 1447
OY 982 TAGGCTGGAACAGAGACTTGGGGGGGATGCTCCAGATGTGGGTGATTTCTGACTGG 1041
DB 1448 TAGGCTGGAACAGAGACTTGGGGGGGATGCTCCAGATGTGGGTGATTTCTGACTGG 1507
OY 1042 GAGGCTATCTGCAAGCTTCCAGACAGGAGCACTCCAGGCTCAGGCTCAGGCTCAGGCT 1101
DB 1508 GAGGCTATCTGCAAGCTTCCAGACAGGAGCACTCCAGGCTCAGGCTCAGGCTCAGGCT 1567

```



Marthnew  
09/18/99/19  
seg. 1D 15  
Page 1

GenCore version 5.1.6  
Copyright (c) 1993 - 2003 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: June 28, 2003, 21:36:27 ; Search time 2479 Seconds

(without alignments)  
16400.411 Million cell updates/sec

Title: US-09-989-919-15

Perfect score: 1397  
Sequence: 1 ggtgctgcacgtaccgga.....aaaaaaaaaaagcgcgtc 1397

Scoring table: IDENTITY NUC  
Gapop 10.0, Gapext 1.0

Searched: 2054640 seqs, 14551402878 residues

Total number of hits satisfying chosen parameters: 4109280

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Database:

GeneBld: 1: gb\_ba: 2: gb\_hcg: 3: gb\_in: 4: gb\_om: 5: gb\_ov: 6: gb\_pat: 7: gb\_ph: 8: gb\_pl: 9: gb\_pr: 10: gb\_ro: 11: gb\_sts: 12: gb\_sy: 13: gb\_un: 14: gb\_vl: 15: em\_ba: 16: em\_fun: 17: em\_hum: 18: em\_in: 19: em\_mu: 20: em\_om: 21: em\_or: 22: em\_ov: 23: em\_pat: 24: em\_ph: 25: em\_pl: 26: em\_ro: 27: em\_sts: 28: em\_un: 29: em\_vl: 30: em\_hcg\_hum: 31: em\_hcg\_inv: 32: em\_hcg\_other: 33: em\_hcg\_mus: 34: em\_hcg\_pin: 35: em\_hcg\_rod: 36: em\_hcg\_mam: 37: em\_hcg\_vrt: 38: em\_sy: 39: em\_higo\_hum: 40: em\_higo\_mus: 41: em\_higo\_other:

Pred. No. is the number of results predicted by chance to have a.

score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

# SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	1397	100.0	1397	6	AX480725
2	1309	93.7	12866	9	AC013480
3	948	67.9	1493	9	BC007901
4	940.4	67.3	2461	6	AX056460
5	797.2	35.6	9436	6	AC112468
6	464.2	33.2	470	6	AX480724
7	420.8	30.1	427	6	AX08491
8	400	28.6	176771	2	AC122406
9	254.4	18.2	462	11	G21938
10	251	18.0	1586	10	BC022157
11	235.8	16.9	1886	6	AX056461
12	56.6	4.1	278726	2	AC099468
13	55.8	4.0	200207	2	AC117061
14	53.2	3.8	133501	2	AC116956
15	52.8	3.8	158007	2	AC121641
16	52.4	3.8	213991	2	AC119579
17	52.2	3.7	86904	2	AC117177
18	51.4	3.7	110000	2	AC096324_0
19	51.4	3.7	110000	2	AC096324_1
20	51	3.7	83391	6	AX458577
21	50.8	3.6	17720	2	AC117075
22	50.8	3.6	125020	9	AF429315
23	50.6	3.6	425	6	AX284646
24	50.6	3.6	16688	6	AX345222
25	50.2	3.6	9716	2	AC115603
26	49.6	3.6	47791	2	AC115593
27	49.6	3.6	47791	2	AC116305
28	49.6	3.6	134541	2	AL627255
29	49.4	3.5	3795	6	AX052773
30	49.2	3.5	268147	2	AC116966
31	49	3.5	78520	2	AP003869
32	49	3.5	125020	2	AF429315
33	49	3.5	149531	2	AP004654
34	49	3.5	160893	2	AC094329
35	49	3.5	177204	2	AC107116
36	48.8	3.5	105470	2	AC116306
37	48.8	3.5	132000	2	AC116976
38	48.8	3.5	256774	2	AC116964
39	48.6	3.5	27291	2	AC115575
40	48.6	3.5	55470	2	AC116989
41	48.6	3.5	171187	2	AC116960
42	48.4	3.5	31343	2	AC115576
43	48.4	3.5	47791	2	AC115593
44	48.4	3.5	47791	2	AC116305
45	48.2	3.5	436	6	AX284728

## ALIGNMENTS

RESULT 1	LOCUS	DEFINITION	ACCESSION	VERSION	KEYWORDS	SOURCE	ORGANISM	REFERENCE	AUTHORS	TITLE
AX480725	AX480725	Sequence 15 from Patent WO0242460.	AX480725	1	GI:22217467	human.	human.	1	Macina, R.A., Recipon, H., Pluta, J., Ghosh, M.G., Sun, Y. and Lai, C.	Compositions and methods relating to colon specific genes and proteins
AX480725	AX480725	1397 bp	AX480725	1	GI:22217467	human.	human.	1	Macina, R.A., Recipon, H., Pluta, J., Ghosh, M.G., Sun, Y. and Lai, C.	Compositions and methods relating to colon specific genes and proteins
AX480725	AX480725	1397 bp	AX480725	1	GI:22217467	human.	human.	1	Macina, R.A., Recipon, H., Pluta, J., Ghosh, M.G., Sun, Y. and Lai, C.	Compositions and methods relating to colon specific genes and proteins

JOURNAL Patent: WO 0242460-A 15 30-MAY-2002;  
Diadexus, Inc. (US)  
Location/Qualifiers  
1.1397  
/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
BASE COUNT 332 a 363 c 398 g 304 t  
ORIGIN

Query Match 100.0%; Score 1397; DB 6; Length 1397;  
Best Local Similarity 100.0%; Pred. No. 0;  
Matches 1397; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

1 GGTGCTGACCTGTACCGGAGCGGAGATCTGAGAACTCCAGGAGAGAGATAC 60  
1 GGTGCTGACCTGTACCGGAGCGGAGATCTGAGAACTCCAGGAGAGAGATAC 60  
61 CGAGTACAGGTATCCAGAGACCAATCCCGAGAGAACTACCGCTCTGGCCATC 120  
61 CGAGTACAGGTATCCAGAGACCAATCCCGAGAGAACTACCGCTCTGGCCATC 120  
121 CTACACACAGGAGAGCTGCTCTTCACTGTTCACTGCTGAGCTGTGATCTG 180  
121 CTACACACAGGAGAGCTGCTCTTCACTGTTCACTGCTGAGCTGTGATCTG 180  
181 TGAGAGCCATGCGGAGCTGCTGCTTGTGATCAACAGAGAGAGAGAGATG 240  
181 TGAGAGCCATGCGGAGCTGCTGCTTGTGATCAACAGAGAGAGAGATG 240  
241 GCGCTGGAGAGAGAGCTTCCAGAGAGATGAGAGAGAGAGAGAGAGATG 300  
241 GCGCTGGAGAGAGAGCTTCCAGAGAGATGAGAGAGAGAGAGAGAGATG 300  
301 TGATCCCATCGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGATG 360  
301 TGATCCCATCGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGATG 360  
361 TCAAGGAT 420  
361 TCAAGGAT 420  
421 TTACTCTACCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 480  
421 TTACTCTACCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 480  
481 AAGTGTCTCTGATCCCAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAT 540  
481 AAGTGTCTCTGATCCCAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAT 540  
541 GGTCTGGAT 600  
541 GGTCTGGAT 600  
601 GCACTGACAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAT 660  
601 GCACTGACAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAT 660  
661 GACAGATGACAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAT 720  
661 GACAGATGACAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAT 720  
721 CTTAGCTATGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAT 780  
721 CTTAGCTATGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAT 780  
781 ACCCTGTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAT 840  
781 ACCCTGTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAT 840  
841 GGGAGGCTTCTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAT 900  
841 GGGAGGCTTCTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAT 900

901 CCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 960  
901 CCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 960  
961 TCGGAT 1020  
961 TCGGAT 1020  
1021 TGTGGGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1080  
1021 TGTGGGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1080  
1081 GCGAGCCAGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAT 1140  
1081 GCGAGCCAGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAT 1140  
1141 GAT 1200  
1141 GAT 1200  
1201 CATTTCTTCAATGAAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAT 1260  
1201 CATTTCTTCAATGAAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAT 1260  
1261 TGAATATCAGAAATCTTTGATCTTCTTGTGATTAATTTGTAATAAT 1320  
1261 TGAATATCAGAAATCTTTGATCTTCTTGTGATTAATTTGTAATAAT 1320  
1321 AAAATTAATTTGATTAATTAATTAATTAATTAATTAATTAATTAAT 1380  
1321 AAAATTAATTTGATTAATTAATTAATTAATTAATTAATTAATTAAT 1380  
1381 AAAATTAATTTGATTAATTAATTAATTAATTAATTAATTAATTAAT 1397  
1381 AAAATTAATTTGATTAATTAATTAATTAATTAATTAATTAATTAAT 1397

RESULT 2  
AC013480  
LOCUS AC013480 128466 bp DNA linear PRI 07-NOV-2001  
DEFINITION Homo sapiens BAC clone RP11-564016 from 2, complete sequence.  
AC013480  
VERSION AC013480.12 GI:16756369  
KEYWORDS HTG.  
SOURCE Homo sapiens.  
ORGANISM Homo sapiens.  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.  
REFERENCE 1 (bases 1 to 128466).  
Toward a complete human genome sequence  
Genome Res. 8 (11), 1097-1108 (1998)  
JOURNAL  
MEDLINE  
PUBMED  
9847074  
REFERENCE 2 (bases 1 to 128466)  
Nguyen, C. and Abbott, A.  
Toward a complete human genome sequence  
Genome Res. 8 (11), 1097-1108 (1998)  
JOURNAL  
TITLE The sequence of Homo sapiens BAC clone RP11-564016  
REFERENCE 3 (bases 1 to 128466)  
Unpublished  
AUTHORS Waterston, R.H.  
JOURNAL  
TITLE Direct Submission  
Submitted (12-NOV-1999) Genome Sequencing Center, Washington  
University School of Medicine, 4444 Forest Park Parkway, St. Louis,  
MO 63108, USA  
REFERENCE 4 (bases 1 to 128466)  
Waterston, R.  
JOURNAL  
TITLE Direct Submission  
Submitted (07-NOV-2001) Department of Genetics, Washington  
University, 4444 Forest Park Avenue, St. Louis, Missouri 63108, USA  
On Nov 7, 2001 this sequence version replaced gi:1339435.  
COMMENT  
Genome Center  
Center: Washington University Genome Sequencing Center  
Center code: WUGSC

Web site: <http://genome.wustl.edu/gsc>  
Contact: sapiens@wustl.wustl.edu  
----- Summary Statistics -----  
Center project name: H\_NH0564016

NOTICE: This sequence may not represent the entire insert of this clone. It may be shorter because we only sequence overlapping clone sections once, or longer because we provide a small overlap between neighboring data submissions.

This sequence was finished as follows unless otherwise noted: all regions were double stranded, sequenced with an alternate chemistry, or covered by high quality data (i.e., phred quality >= 30); an attempt was made to resolve all sequencing problems, such as compressions and repeats; all regions were covered by sequence from more than one subclone; and the assembly was confirmed by restriction digest.

**MAPPING INFORMATION:**  
Mapping information for this clone was provided by Dr. John D. McPherson, Department of Genetics, Washington University, St. Louis MO. For additional information about the map position of this sequence, see <http://genome.wustl.edu/gsc>

**SOURCE INFORMATION:**  
The RPc1-11 human BAC library was made from the blood of one male donor, as described by Osoegwa, K., Moon, P.Y., Zhao, B., Fritngen, E., Tatenno, M., Caranese, J.J. and de Jong, P.J. (1998). An improved approach for construction of bacterial artificial chromosome libraries. *Genomics* 51:1-8. The clone may be obtained either from Research Genetics, Inc. (<http://www.regen.com>) or Pieter de Jong and coworkers at the Roswell Park Cancer Institute (<http://bacpac.med.buffalo.edu>)

NEIGHBORING SEQUENCE INFORMATION:  
The clone sequenced to the left is AC068552; the clone sequenced to the right is C02-0096A9. Actual start of this clone is at base position 1 of RP11-564016.

The sequence from base position 103537 to 103654 and from base position 103940 to 104887 is derived from PCR of BAC DNA.

The fidelity of the sequence from base position 48371 to 48772 cannot be guaranteed due to an unresolved imperfect dinucleotide repeat.

## FEATURES

```

source
location/Qualifiers
1..128466
/organism="Homo sapiens"
/db_xref="taxon:9606"
/chromosome="2"
/map="2"
/clone="RP11-564O16"
/clone_1fb="RPCL-11"
291..752
/rpt_family="Alu"
/rpt_family="L2"
1484..1617
/rpt_family="L2"
1669..1826
/rpt_family="Alu"
1864..2412
/rpt_family="MALR"
2413..2591
/rpt_family="MIR"
2689..2833
/rpt_family="L2"
3035..3082
/rpt_family="(T)n"
3085..3392
/rpt_family="Alu"
3530..3643
/rpt_family="CRI"
3644..3717
repeat_region

```

```

repeat_region /rpt_family="Maximer"
3718 .4183
/rpt_family="CRI"
repeat_region /rpt_family="CRI"
6123 .6293
/rpt_family="MIR"
repeat_region /rpt_family="MIR"
6845 .6881
/rpt_family="AT_rich"
repeat_region /rpt_family="AT_rich"
6911 .6963
/rpt_family="AT_rich"
7910 .8008
/rpt_family="MIR"
repeat_region /rpt_family="MIR"
8009 .8299
/rpt_family="Alu"
repeat_region /rpt_family="Alu"
8300 .8326
/rpt_family="MIR"
repeat_region /rpt_family="Alu"
8327 .8599
/rpt_family="Alu"
repeat_region /rpt_family="MIR"
8600 .8668
/rpt_family="MIR"
repeat_region /rpt_family="MIR"
9463 .9553
/rpt_family="TTCA)n"
repeat_region /rpt_family="TTCA)n"
9890 .9923
/rpt_family="AT_rich"
repeat_region /rpt_family="AT_rich"
10066 .10473
/rpt_family="L2"
repeat_region /rpt_family="L2"
10474 .10520
/rpt_family="CA)n"
repeat_region /rpt_family="CA)n"
10521 .10746
/rpt_family="L2"
repeat_region /rpt_family="L2"
12200 .12385
/rpt_family="MIR"
repeat_region /rpt_family="MIR"
12455 .12489
/rpt_family="CACC)n"
repeat_region /rpt_family="CACC)n"
13410 .13773
/rpt_family="MALR"
repeat_region /rpt_family="MALR"
14076 .14317
/rpt_family="MIR"
repeat_region /rpt_family="MIR"
15470 .15592
/rpt_family="MIR"
repeat_region /rpt_family="MIR"
15599 .15967
/rpt_family="MALR"
repeat_region /rpt_family="MALR"
16227 .16617
/rpt_family="MALR"
misc_feature /note="similar to EST AA913068 (NID:G3052460) ol34e06.s1"
16517 .16693
/rpt_family="L1"
repeat_region /rpt_family="L1"
16717 .16788
/rpt_family="L1"
repeat_region /rpt_family="L1"
16985 .17009
/rpt_family="AT_rich"
repeat_region /rpt_family="AT_rich"
18369 .18509
/rpt_family="MER103"
repeat_region /rpt_family="MER103"
18594 .18644
/rpt_family="TG)n"
repeat_region /rpt_family="TG)n"
19032 .19193
/rpt_family="MER1_type"
repeat_region /rpt_family="MER1_type"
19359 .19668
/rpt_family="L2"
repeat_region /rpt_family="L2"
22454 .22498
/rpt_family="GA-rich"
repeat_region /rpt_family="GA-rich"
25334 .25544
/rpt_family="Alu"
repeat_region /rpt_family="Alu"
25647 .25768
/rpt_family="MIR"
repeat_region /rpt_family="MIR"
25770 .25888
/rpt_family="L2"
repeat_region /rpt_family="L2"
25893 .26193
/rpt_family="Alu"
misc_feature /rpt_family="Alu"
25910 .26358
/rpt_family="Alu"
misc_feature /note="similar to EST AA578847 (NID:G2357031) nk66a09.s1"
26366 .26432
/rpt_family="MIR"
repeat_region /rpt_family="MIR"
26502 .26568
/rpt_family="MIR"

```

Query Match	93.7%	Score 1309;	DB 9;	Length 128466;
Best Local Similarity	99.6%	Pred. No. 0;		
Matches 1312; Conservative	0;	Mismatches 5;	Indels 0;	Gaps 0;

QY	779	CAACCTGAGGCTGTAAAGCAAGCTCAGGCTAGTCTCCCACTGGGGGCTGTGCCCTCC	838
Db	114515	CAACCTGTGGGCTGTAAAGCAAGCTCAGGCTAGTCTCCCACTGGGGGCTGTGCCCTCC	114574
QY	839	CTGGACAGGTTCCGTGGGACAGCCCATCACTGTGTTCAATAGTGTGAATATGTAGCTAA	898
Db	114575	CTGGACAGGTTCCGTGGGACAGCCCATCACTGTGTTCAATAGTGTGAATATGTAGCTAA	114634
QY	899	GCCCTGCTGCTGCTGCTGCAATGCCACAGACGGGTGGGGGCTGCGTGGGACAATC	958
Db	114635	GCCCTGCTGCTGCTGCTGCAATGCCACAGACGGGTGGGGGCTGCGTGGGACAATC	114694
QY	959	CATGTGAGTGTCTCAGCTTAGTGTGGAACAGAGACTGTGGGGGGATGTGCTCAG	1018
Db	114695	CATGTGAGTGTCTCAGCTTAGTGTGGAACAGAGACTGTGGGGGGATGTGCTCAG	114754
QY	1019	GATGTGGTGATTTCTGTACCTGTGGGAGGCTATCTTGACTCCGACAGGGGACACTCC	1078
Db	114755	GATGTGGTGATTTCTGTACCTGTGGGAGGCTATCTTGACTCCGACAGGGGACACTCC	114814
QY	1079	AGGCCAGCCACAGGGGTACAGGGGCAGAGGTGCACACTCAGCATGAGCCAAAGCTGGGGTC	1138
Db	114815	AGGCCAGCCACAGGGGTACAGGGGCAGAGGTGCACACTCAGCATGAGCCAAAGCTGGGGTC	114874
QY	1139	AGGAGACAGGTGTGGTTGAGCCAGAGACTCTGGGGGTGGGGCCGGGCTTTCTGCG	1198
Db	114875	AGGAGACAGGTGTGGTTGAGCCAGAGACTCTGGGGGTGGGGCCGGGCTTTCTGCG	114934
QY	1199	CTCATTTGCTTTCAATGAAGACCTCAAGACAGCCAAACAGGCTTTCCCTCTCTGGA	1258
Db	114935	CTCATTTGCTTTCAATGAAGACCTCAAGACAGCCAAACAGGCTTTCCCTCTCTGGA	114994
QY	1259	GTTTGATATCCAGAACTTTTGTACTCTTGTGGTTAAATGTGTTATTTTGTAAAAA	1318
Db	114995	GTTTGATATCCAGAACTTTTGTACTCTTGTGGTTAAATGTGTTATTTTGTAAAAA	115054
QY	1319	ATAAAATPAAATAGTTAATPAAATGATGTTTCCACAGAAACTCTTCCCTAAAAAA	1375
Db	115055	ATAAAATPAAATAGTTAATPAAATGATGTTTCCACAGAAACTCTTCCCTAAAAAA	115111

[illegible]

Lim, M., Maduro, Q. L., Masiello, C., Mastrian, S. D., McCloskey, J. C., McDowell, J., Pearson, R., Snyder, B., Stantrop, S., Thomas, P. J., Tongson, E. B., Touchman, J. W., Teurigne, C., Vogt, J. L., Walker, M. A., Zhang, L. H. and Green, E. D.

Clone distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/HLN at: <http://image.llnl.gov>  
Series: IRL Plate: 20 Row: 3 Column: 15  
This clone was selected for full length sequencing because it passed the following selection criteria: Hexamer frequency ORF analysis.

## FEATURES

Location/Qualifiers

1. 1493  
/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
/clone="IMAGE:4139786"  
/tissue\_type="Muscle, rhabdomyosarcoma"  
/clone\_id="N1H\_MGC\_17"  
/lab\_host="DH10B-R"  
/note="Vector: POTB7"  
<1. 625  
/codon\_start=2  
/product="Unknown (protein for IMAGE:4139786)"  
/protein\_id="AAH07901.1"  
/db\_xref="GI:14043928"  
/translation="ARGELKVTLDARVEEPACSTDCILEFPARNFTLPCSAQGM  
CEGMEKRLVNAVRFPPYLPBPSAPSLDLSIVATGELAGVETLAELEKY  
LHLYRGOYLQNSTASSSTHYOCIPDSTIPQBYRWPSYHSHSCLSLSPNLEAVDV  
CESHACRAFPVNTMTWRQLVFRKTMQVDPNKTIVKASG"

## CDS

BASE COUNT 360 a 401 c 422 g 310 t  
ORIGIN

Query Match 67.9%; Score 948; DB 9; Length 1493;  
Best Local Similarity 84.3%; Pred. No. 7.1e-238;

Matches 112; Conservative 0; Mismatches 10; Indels 208; Gaps 1;

1 GGTGTCGACCTGTACCGAGCGGAGATGTGCAAACTCCACGCGACAGCAGTAC 60  
304 GGTGTCGACCTGTACCGAGCGGAGATGTGCAAACTCCACGCGACAGCAGTAC 363  
61 CGAGTACCACTGTATCCAGACAGCAGCATCCCCAGAAAGATACCGCTGCGCCATC 120  
364 CGAGTACCACTGTATCCAGACAGCAGCATCCCCAGAAAGATACCGCTGCGCCATC 423  
121 CTACCAACACGGAGGAGCTCTCTTCAAGTTCACACCTGCGAGGCTGTGATGTC 180  
424 CTACCAACACGGAGGAGCTCTCTTCAAGTTCACACCTGCGAGGCTGTGATGTC 483  
181 TGAGAGCCATGCCAGTGTGCGGCTTTGTGTCAACACAGACCACTGACAGATGA 240  
484 TGAGAGCCATGCCAGTGTGCGGCTTTGTGTCAACACAGACCACTGACAGATGA 536  
241 GCCAGTGGAGAAACCTTTCAGAGGAGATGGACAGACCTCTGTGAGGTTGATGATG 300  
537 ----- 536  
301 TGATCCCCCATCGGAAGTCAAGAGGGGTGCTGAGTGATGAGAGAGATATACGTCT 360  
537 ----- 536  
361 TCAGAGCAGTCAATTAGGAGAGATGCTTGTCCACAGAAAGAAACATCCAGCCCTG 420  
537 ----- 536  
421 TTACTCTCACTCTGCCCCCAGGTGCGAGCTGCTTTTCAAGACTGATGAGCC 480  
537 ----- 536  
481 AAGTGTCCCTGATCCCAACAGACACATATGTAAGGCTCTGCTGACCTATCTGAG 540  
576 AAGTGTCCCTGATCCCAACAGACACATATGTAAGGCTCTGCTGACCTATCTGAG 635  
541 GGCTCGGCTGACCAAGTGTACTATCTCAGCAGCTGGGCTTGGCTGTGAGGAGATGATT 600

636 GGCTCGGCTGACCAAGTGTACTATCTCAGCAGCTGGGCTTGGCTGTGAGGAGTACTT 695  
601 GGAAGTGGACGACCTGATGTCACCTGCGGAAACCTGACAGCAAAAGTACATCCAGACA 660  
696 GGAAGTGGACGACCTGATGTCACCTGCGGAAACCTGACAGCAAAAGTACATCCAGACA 755  
661 GACAGATGTGACCAAGACCAAAAGTGAATATGCAATGTTAAATGATGATTTACCG 720  
756 GACAGATGTGACCAAGACCAAAAGTGAATATGCAATGTTAAATGATGATTTACCG 815  
721 CCTAGCTATGGAATCTGCTGCTCTTATGTCACAGAAATCATGCGGAGTATGACTGCTTCA 780  
816 CCTAGCTATGGAATCTGCTGCTCTTATGTCACAGAAATCATGCGGAGTATGACTGCTTCA 875  
781 ACCCTGTGGGCTGTATGACAAAGCTCAAGGCTGCTCCCACTGGGGGCTGTGCCCTCCCT 840  
876 ACCCTGTGGGCTGTATGACAAAGCTCAAGGCTGCTCCCACTGGGGGCTGTGCCCTCCCT 935  
841 GGAGACGCTTCCGTGGGACCCCATCACTGTGTTCATATGATGATGATGATGATGATGATG 900  
936 GGAGACGCTTCCGTGGGACCCCATCACTGTGTTCATATGATGATGATGATGATGATGATG 995  
901 CCTGTGCTGCTGCTGCTGACATGCAACAGACAGCGGCTGGGCTGGGCGCAATGCA 960  
996 CCTGTGCTGCTGCTGCTGACATGCAACAGACAGCGGCTGGGCTGGGCGCAATGCA 1055  
961 TGTGTGAGTGTCTCTCAGCTTATGCTGTGACAGAGATCTGGCGGGGATGCTCCAGTA 1020  
1056 TGTGTGAGTGTCTCTCAGCTTATGCTGTGACAGAGATCTGGCGGGGATGCTCCAGTA 1115  
1021 TGTGTGAGTGTCTCTCAGCTTATGCTGTGACCTCCGACAGGGGACATCTCCAG 1080  
1116 TGTGTGAGTGTCTCTCAGCTTATGCTGTGACCTCCGACAGGGGACATCTCCAG 1175  
1081 GCGAGCCAGGGGCTCAAGGAGTGTGACACCTTCAGATGAGCCAAAGATCTGGGCTCAG 1140  
1176 GCGAGCCAGGGGCTCAAGGAGTGTGACACCTTCAGATGAGCCAAAGATCTGGGCTCAG 1235  
1141 GAGACAGTGTGCTTGTGACCCAGACCTTGGGGGCTGGGCTGGGCTTGTGCTCT 1200  
1236 GAGACAGTGTGCTTGTGACCCAGACCTTGGGGGCTGGGCTGGGCTTGTGCTCT 1295  
1201 CATTGCTTCAATGAAAGCTCAAGAGCAGCCAAACAGAGCTTCCCTCTCTGAGT 1260  
1296 CATTGCTTCAATGAAAGCTCAAGAGCAGCCAAACAGAGCTTCCCTCTCTGAGT 1355  
1261 TTGAATATCCAGAACTTTTGTACTTCTTGTGTTAAATGTTATTTTGTAAAAAT 1320  
1356 TTGAATATCCAGAACTTTTGTACTTCTTGTGTTAAATGTTATTTTGTAAAAAT 1415  
1321 AAAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 1380  
1416 AAAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 1475  
1381 AAAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 1390  
1476 AAAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 1485

RESULT 4  
AX056460 2461 bp DNA linear PAT 13-JUN-2001

DEFINITION Sequence 104 from Patent WO0073469.

AX056460  
VERSION AX056460.1 GI:12229167

KEYWORDS

SOURCE human.

ORGANISM Homo sapiens

REFERENCE  
AUTHORS Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo. Plowman, G. D., Martinez, R., Whyte, D. and Sudersanam, S.



TITLE Protein Kinases  
JOURNAL Patent: WO 0073469-A 104 07-DEC-2000;  
Sugen, Inc. (US)  
FEATURES Location/Qualifiers  
Source 1. 2461  
/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
BASE COUNT 426 a 771 c 811 g 453 t  
ORIGIN

Query Match 67.3%; Score 940.4; DB 6; Length 2461;  
Best Local Similarity 84.7%; Pred. No. 7.6e-236;  
Matches 1159; Conservative 0; Mismatches 1; Indels 208; Gaps 1;

1 GGTGTGACCTGATGACGAGGCGGAGTATCTGACAACTCCAGCGGACAGCAGTAC 60  
1302 GGTGTGACCTGATGACGAGGCGGAGTATCTGACAACTCCAGCGGACAGCAGTAC 1361

61 CGAGTACCACTGATGATCCAGACAGACACATCCCGAGGAGACTACCGCTGCTGGCATC 120  
1362 CGAGTACCACTGATGATCCAGACAGACACATCCCGAGGAGACTACCGCTGCTGGCATC 1421

121 CTACACACAGCGGAGCTGCTCTTTCACTGATGATGATGATGATGATGATGATGAT 180  
1422 CTACACACAGCGGAGCTGCTCTTTCACTGATGATGATGATGATGATGATGATGAT 1481

181 TGAGAGCAGTATGAGTATGAGTATGAGTATGAGTATGAGTATGAGTATGAGTATGAGT 240  
1482 TGAGAGCAGTATGAGTATGAGTATGAGTATGAGTATGAGTATGAGTATGAGTATGAGT 1534

241 GCCAGTGGGAGAACCCCTTCCAGAGGAGATGGCAGAGCTCTGAGGTTGATAGATAG 300  
1535 ----- 1534

301 TGATCCCCCATCGAAGTCAAGAGGGGCTGCTGAGGTATGAGAGAGATATACGTCT 360  
1535 ----- 1534

361 TCAAGGAGTCAATTAAGGAGATGCTTGTGCTCCAGAAAGAAACATCCAGCCCTG 420  
1535 ----- 1534

421 TTACCTCTACCTCTGCCCCCAGGTGGGAGCTGCTTTTCAAGACTGATGAGGCC 480  
1535 ----- 1534

481 AAGTGTCCCTGATATCCCAAGACCAATATGAGGAGCTCTGCTGATCTATCTGAG 540  
1574 AAGTGTCCCTGATATCCCAAGACCAATATGAGGAGCTCTGCTGATCTATCTGAG 1633

541 GGTGTGACCTGATGACGAGGCGGAGTATCTGACAACTCCAGCGGACAGCAGTAC 600  
1634 GGTGTGACCTGATGACGAGGCGGAGTATCTGACAACTCCAGCGGACAGCAGTAC 1693

601 GCACTGGCAGCAGCTGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 660  
1694 GCACTGGCAGCAGCTGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1753

661 GACAGATGTGACAGAGGCAAAAGGCAATATGAGGAGCTCTGCTGATCTATCTGAG 720  
1754 GACAGATGTGACAGAGGCAAAAGGCAATATGAGGAGCTCTGCTGATCTATCTGAG 1813

721 CCTAGCTATGAGCTGCTGCTCTTATGATGAGGAGTATGATGATGATGATGATGATGAT 780  
1814 CCTAGCTATGAGCTGCTGCTCTTATGATGAGGAGTATGATGATGATGATGATGATGAT 1873

781 ACCCTGTGAGCTGATGAGCTGATGATGATGATGATGATGATGATGATGATGATGATGAT 840  
1874 ACCCTGTGAGCTGATGAGCTGATGATGATGATGATGATGATGATGATGATGATGATGAT 1933

841 GGGAGCGTTCCTGGGAGCGGACCATCATGTTTCAATGATGAGTATGATGATGATGATGAT 900  
1934 GGGAGCGTTCCTGGGAGCGGACCATCATGTTTCAATGATGAGTATGATGATGATGATGAT 1993

QY 901 CCTGTGCTGCTGCTGCTGACATCCACAGCAGCGGCTGGGCTGCTGGGACAAATCCA 960  
DB 1994 CCTGTGCTGCTGCTGCTGACATCCACAGCAGCGGCTGGGCTGCTGGGACAAATCCA 2053

QY 961 TCGGAGGATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1020  
DB 2054 TCGGAGGATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 2113

QY 1021 TGTGGGATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1080  
DB 2114 TGTGGGATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 2173

QY 1081 GCCAGCCAGGAGGCTGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 1140  
DB 2174 GCCAGCCAGGAGGCTGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 2233

QY 1141 GGAGCAGGATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1200  
DB 2234 GGAGCAGGATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 2293

QY 1201 CATTGCTTTCATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1260  
DB 2294 CATTGCTTTCATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 2353

QY 1261 TTGAATATCCAGATCTTTTGTATCTTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1320  
DB 2354 TTGAATATCCAGATCTTTTGTATCTTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 2413

QY 1321 AAAATAAATTTAGTAAATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1368  
DB 2414 AAAATAAATTTAGTAAATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 2461

RESULT 5  
LOCUS AC112468  
DEFINITION Rattus norvegicus clone CH230-12986, \*\*\* SEQUENCING IN PROGRESS  
AC112468 94366 bp DNA linear HNG 17-JUL-2002  
AC112468:2 GI:21745511  
KEYWORDS HTG; HTGS PHASE1.  
SOURCE Norway rat.  
ORGANISM Rattus norvegicus  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;  
Rattus.

## REFERENCE

1 (bases 1 to 94366)  
Muzny,D.M., Adams,C., Adio-Oduola,B., Ali-ouman,F.R., Allen,C.,  
Alsbrooks,S.L., Amaralunge,H.C., Are,J.R., Ayele,M., Banks,T.,  
Barbatta,J., Benton,J., Bimage,K., Blankenburg,K., Bonnin,D.,  
Bouck,D., Bowie,S., Blevins,M., Brown,E., Brown,M., Bryant,N.P.,  
Buhay,C., Burch,P., Burkett,C., Burrell,K.L., Byrd,N.C.,  
Carroll,T.F., Carter,M., Cavazos,S.R., Chacko,J., Chavez,D.,  
Chen,G., Chen,R., Chen,Z., Chowdhury,I., Christopoulos,C.,  
Cleveland,C.D., Cox,C., Coyle,M.D., Dathorne,S.R., David,R.,  
Davila,M.L., Davis,C., Day-Carroll,L., Dedrich,D.A.,  
Delaney,K.R., Delgado,O., Dem,A.V., Ding,Y., Dinh,H.H.,  
Douthwaite,K.J., Draper,H., Dugan-Rocha,S., Durbin,K.J.,  
Earnhart,C., Edgar,D., Edwards,C.C., Elhaj,C., Escoto,M.,  
Falls,T., Ferraguto,D., Flagg,N., Ford,J., Foster,P., Frantz,P.,  
Gabriel,A., Gao,J., Garcia,A., Garner,T., Garza,N., Gill,R.,  
Gorell,J.H., Guevara,W., Gunaratne,P., Hale,S., Hamilton,K.,  
Harris,C., Harris,K., Hart,M., Hayla,P., Hawes,A., Hernandez,J.,  
Hernandez,O., Hodgson,A., Hughes,M., Hollaway,C., Hollins,B.,  
Homi,F., Howard,S., Huber,J., Huliy,K., Hume,J., Jackson,L.E.,  
Jacobson,B., Jia,Y., Johnson,R., Jolivet,S., Joudah,S.,  
Karissoun,E., Kelly,S., Khan,U., King,L., Korvah,J., Kovar,C.,  
Kratovic,J., Kureshi,A., Landry,N., Leal,B., Lewis,L.C., Lewis,L.,  
Li,J., Li,Z., Lichtarge,O., Lieu,C., Liu,J., Liu,M., Louissege,H.,  
Lozdo,R.D., Lu,X., Lucier,A., Lucier,R., Luna,R., Ma,J.,  
Maneshwari,M., Mapua,P., Martin,R., Martindale,A., Martinez,E.,  
Massey,E., Mawhiney,E., Mcleod,M.P., Meador,M., Mei,G., Metzker,M.,



TITLE  
JOURNAL  
REFERENCE  
AUTHORS  
JOURNAL  
REFERENCE  
AUTHORS  
JOURNAL  
COMMENT

Miner, G., Miner, Z., Mitchell, T., Mohabbat, K., Morgan, M., Morris, S., Moser, M., Neal, D., Newton, J., Newton, N., Nguyen, A., Nguyen, N., Nguyen, N., Nguyen, N., Nwokwu, S., Ogun, M., Okunola, G., Oragunye, N., Oviedo, R., Pace, A., Payton, B., Peery, J., Perez, L., Peters, L., Pickens, R., Primus, E., Pul, L., Quiles, M., Ren, Y., Rives, M., Rojas, A., Rojudoan, I., Rolfe, M., Ruiz, S., Severy, G., Scheraga, S., Scott, G., Shen, H., Shoshitashvili, N., Sisson, I., Sutton, A., Svatek, A., Tabor, P., Tamerisa, A., Thomas, N., Thomas, S., Tansley, J., Taylor, C., Taylor, T., Telford, B., Thomas, R., Wang, S., Wang, S., Ward-Moore, S., Warren, R., Washington, C., Watlington, S., Williams, G., Williamson, A., Wleczky, R., Wooden, S., Worley, K., Wu, C., Wu, Y., Wu, Y.F., Zhou, J., Zorrilla, S., Nelson, D., Weinstein, G., and Gibbs, R.

Unpublished  
2 (bases 1 to 94366)  
Worley, K.C.  
Direct Submission  
Submitted (21-FEB-2002) Human Genome Sequencing Center, Department of Molecular and Human Genetics, Baylor College of Medicine, One Baylor Plaza, Houston, TX 77030, USA  
3 (bases 1 to 94366)  
Worley, K.C.  
Direct Submission  
Submitted (17-JUL-2002) Human Genome Sequencing Center, Department of Molecular and Human Genetics, Baylor College of Medicine, One Baylor Plaza, Houston, TX 77030, USA  
On Jul 14, 2002 this sequence version replaced gi:18850446.

Genome Center  
Center: Baylor College of Medicine  
Center code: BCM  
Web site: <http://www.hgsc.bcm.tmc.edu/>  
Contact: hgsc-help@bcm.tmc.edu  
Project Information  
Center project name: GS00  
Center clone name: CH230-12986  
Summary Statistics  
Sequencing vector: Plasmid:  
Chemistry: Dye-terminator Big Dye 100% of reads  
Assembly program: Phrap; version 0.990329  
Consensus quality: 53937 bases at least Q40  
Consensus quality: 56592 bases at least Q30  
Consensus quality: 58737 bases at least Q20

NOTE: Estimated insert size may differ from sequence length  
(see [http://www.hgsc.bcm.tmc.edu/docs/Genbank\\_draft\\_data.html](http://www.hgsc.bcm.tmc.edu/docs/Genbank_draft_data.html)).  
NOTE: This is a 'working draft' sequence. It currently  
\* consists of 44 contigs. The true order of the pieces  
\* is not known and their order in this sequence record is  
\* arbitrary. Gaps between the contigs are represented as  
\* runs of N, but the exact sizes of the gaps are unknown.  
\* This record will be updated with the finished sequence  
\* as soon as it is available and the accession number will  
\* be preserved.

1 1066: contig of 1066 bp in length  
1067 1166: gap of unknown length  
1167 2390: contig of 1224 bp in length  
2391 2490: gap of unknown length  
2491 3708: contig of 1218 bp in length  
3709 3808: gap of unknown length  
3809 5425: contig of 1617 bp in length  
5426 5525: gap of unknown length  
5526 6599: contig of 1074 bp in length  
6600 6699: gap of unknown length  
6700 8207: contig of 1508 bp in length  
8208 8307: gap of unknown length  
8308 9744: contig of 1437 bp in length  
9745 9844: gap of unknown length  
9845 11023: contig of 1179 bp in length  
11024 11123: gap of unknown length  
11124 12519: contig of 1396 bp in length

FEATURES  
source

12520 12619: gap of unknown length  
12620 13673: contig of 1054 bp in length  
13674 13773: gap of unknown length  
13774 15620: contig of 1847 bp in length  
15621 15720: gap of unknown length  
15721 17466: contig of 1746 bp in length  
17467 17566: gap of unknown length  
17567 19136: contig of 1570 bp in length  
19137 19236: gap of unknown length  
19237 20973: contig of 1737 bp in length  
20974 21073: gap of unknown length  
21074 22091: contig of 1018 bp in length  
22092 22191: gap of unknown length  
22192 23778: contig of 1581 bp in length  
23779 25293: gap of unknown length  
25294 25394: contig of 1415 bp in length  
25395 26984: gap of unknown length  
26985 27083: contig of 1590 bp in length  
27084 29408: gap of unknown length  
29409 31243: contig of 2325 bp in length  
31244 31343: gap of unknown length  
31344 33604: contig of 2261 bp in length  
33605 33704: gap of unknown length  
33705 35038: contig of 1334 bp in length  
35039 35138: gap of unknown length  
35139 36399: contig of 1261 bp in length  
36400 36499: gap of unknown length  
36500 37623: contig of 1124 bp in length  
37624 40099: gap of unknown length  
40100 40199: contig of 2376 bp in length  
40200 42674: gap of unknown length  
42675 42774: contig of 2475 bp in length  
42775 45617: gap of unknown length  
45618 45717: contig of 2843 bp in length  
45718 47425: gap of unknown length  
47426 47525: contig of 1708 bp in length  
47526 48809: gap of unknown length  
48810 48909: contig of 1264 bp in length  
48910 50692: gap of unknown length  
50693 50792: contig of 1783 bp in length  
50793 52171: gap of unknown length  
52172 52271: contig of 1379 bp in length  
52272 53845: gap of unknown length  
53846 53945: contig of 1574 bp in length  
53946 56501: gap of unknown length  
56502 56601: contig of 2556 bp in length  
56602 59622: gap of unknown length  
59623 59722: contig of 3021 bp in length  
59723 62615: gap of unknown length  
62616 62715: contig of 2893 bp in length  
62716 65623: gap of unknown length  
65624 65723: contig of 2908 bp in length  
65724 67513: gap of unknown length  
67514 70311: contig of 1790 bp in length  
70312 70411: gap of unknown length  
70412 72769: contig of 2638 bp in length  
72770 72869: gap of unknown length  
72870 76193: contig of 3324 bp in length  
76194 76293: gap of unknown length  
76294 79434: contig of 3141 bp in length  
79435 79534: gap of unknown length  
79535 84177: contig of 4643 bp in length  
84178 84277: gap of unknown length  
84278 88578: contig of 4301 bp in length  
88579 94366: gap of unknown length  
94367 12519: contig of 5688 bp in length

Location/Qualifiers  
1. 94366  
/organism="Rattus norvegicus"

BASE COUNT	22805	a	22369	c	21709	g	23135	t	4348	other
ORIGIN										

Query	Match	Best Local Similarity	Score	DB 2:	Length
Matches	928;	Conservative	0;	Mismatches	303;
				Indels	89;
				Gaps	16;
QY	60	CCGAGTATACCAAGTGTATATCCAGACAGACCACTCCCCAGAGAACTACCGCTGTGGCAT	119		
DB	23973	CAGAGTATACCAAGGCAATCCCGGACAGTACCATACACAGAGAGATTATCGCTGTGGCCCT	24033		
QY	120	CCTAACCAACGAGGAGCTGCTCTTCAAGTTCACCTTGGCTGAGGCTGTGGATGCTT	179		
DB	24033	CTTACCAACATGTGTGCTGCTCTCTCTGTCTGTGTTCACCTTGGCTGAGGCTGTGTGATGCT	24092		
QY	180	GTGAGAGCCATGCCACAGTGTGGGCTTTGTGTTCACCAACAGACCACTGGACAGGTG	239		
DB	24093	GTGAGAGCCATGTTCAATGTGTGTGCTTTGTGTTCACCAACAGACCACTGGACAGGTG	24155		
QY	240	AGCAGATGGGAGAAAGCCCTCCACAGGAGATGGACAGACCTCTTGTGAGGTGTATGATA	299		
DB	24153	AGCCAGAGGACAGAGATCTTCAAAAGAG-TGACTGTACTCTTCCAAAGATTTTGA---	24208		
QY	300	GTGATCTCCCATCGAAGTGTACAGAGGAGGTGTCTGATGTATGAGAGAGATTAAGTGT	358		
DB	24209	GGGCCCCCTTCAGAAAGCCCGAGGACAGAGATGACAGAGAGATTTAATATTATA	24268		
QY	359	-----CTTCAAGCACTCAATTTAGGAGATGTCTTGTCTCCAGAAAGAAACATCCA	414		
DB	24269	TAACTTCAAGCACTCACTTAA-TAGAGTAACTTAACTTCAAGAAAGAAATTAACGT	24326		
QY	415	GGCCTGTTAACCTCTCACTCTGCCCCAGGTGGGACGTGTCTTTTCAAGACTGAT	474		
DB	24327	ACTTATATACTTTATCTGTG-CTCTAGTGTGGAAGCTGTGCTTTTAAAGACTGAT	24385		
QY	475	GGAGCAAGTGTCTCCATATCCCAAAAGACACATATGTGAAGGCTCTGGCTGACCTA	534		
DB	24386	GGAAACCAAGTGTCTCCATATCCCAAAAGACACATATGTGAAGGCTCTGGCTGACCTA	24445		
QY	535	TCTGAGGGCTGGCTGTACCAAGCTACTATCTCAGCAGCTGGGCTTGCTGTGGAGGAG	594		
DB	24446	TTGGCGGGCTCG-----CTACACAGCTGGGTTTGTCTGTCTGACAGGGG	24487		
QY	595	TGACTGTGACGTGGCAGACATGTACGTACTGTGGAAACCTCTGACAAAGCTTAATCC	654		
DB	24488	TGACTGTGATG-----CACTCGGGAACCTCTGACAAAGCTTAATCC	24532		
QY	655	CAGACAGACATGTGACAGAGCAAACTGCAATATCCCAATATTTAAATGTGATTT	714		
DB	24533	CAGAGCAACATGTGTGACAGAGCAAACTGCAATATTTAAATGTGATTT	24592		
QY	715	TACAGGCTGATGTGAGATCTGTGGCTCTTGTATCCAGATCA-----TGCGGTATGAC	770		
DB	24593	TGCGAGCTTTGGCTTAAAGCTGTCTCTCCAGCTGAGCTGATCTGTGTGGATTTGAT	24652		
QY	771	TGCTCTTCAACCTGTGGGCTGTAAAGCAAGCTCAGGCTAATCTCCACTGGGGGCTGT	830		
DB	24653	TGCTCTTCTGAGCTGTGAGCAGGGCTCAGGC-----T	24685		
QY	831	GGCCCTCCCGGAGCGTTCCGTGGGACGCCCATCATCTGTGTTCAATAGTGAAGATG	890		
DB	24686	GGTCTTAACTGGGATGTGCTCCAGAGGACGCCCATTAATCTGATTC-ACAGTTTGAATG	24744		
QY	891	TAGCTAAAGCCCTGTGGCTGTGTGACATGACACAGCAGGGGCTGGGGGCTGGCTGG	950		
DB	24745	TAGCCAGAA--CACTGTGCTGCTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT	24802		
QY	951	GGAGCAATCATGTGTGAGTGTCTCTCACTTAAGTGTGACAGAGACTTGTGGCGGAGA	1010		
DB	24803	GGAGCAATCATGTGTGAGTGTCTCTCACTTAAGTGTGACAGAGACTTGTGGCGGAGA	24859		

QY	1011	TGCTCAGAGATGTGGGTGATTTCTGTACCTGGGAGGCTATCTCTGAACCTCCGACAGGG	1070
Db	24860	TGCTTCAAGATGTATAGGAGGACAT--CCGTGGGAAGGCTCTCTGGACCTTCTCCCTACGG	24911
QY	1071	AACATCCAGGACCGACGCCAGGGGTCAAGGGGCAGAGGTCACACCTCAGCATGAGCCAA	1130
Db	24917	TTGGTCCCTCAGCCAGCCC-CAGGTCAATGGGACAGCAGTGCACACATCATATGAGCCAA	24977
QY	1131	CTGGGGTCAGGAGACAGGTGTGTTTGAAGCCAGGACCTTGGGCGGGGTGGGGCCGGGCG	1190
Db	24976	CTAAGAGTGAGAGACAGATTCGACATTTGAGCCAGAGACCTGGGGGTGGGGTGGGGTGGGGG	25035
QY	1191	CTTTCGAGCTCATTTGGTTCATATGAAAGCTCAAGCGAGCCAAACCGAGCTTTCGCC	1250
Db	25036	CTTTCCTCCTCATTTGTTTCATGTAAAGCCAAAGCGAGCCGAAACCGAGCTTTCGCC	25095
QY	1251	TTCTCTCAGTTGAATTCACAGATCTTTTGTACTCTCTGTGGTTAAATTGGTTATTTT	1310
Db	25096	-TCTTGAGGTGTGATATTCAGAGCTTTTGTACTCTCTGTTCATGAAATTTGTTATTTT	25154
QY	1311	TGTAAAAAATTAATAATAATTAATTAATAATAATGATGTTTCACAGCAACCTTCGCCAA	1370
Db	25155	TGTAAAAAATAAT--AATCATTAATAATAATGATGTTTGTGCAAAATACTCCCTCA	25212

RESULT 6					
LOCUS	AX480724				
DEFINITION	Sequence 14 from Patent WO0242460.	470 bp	DNA	linear	PAT 12-AUG-2002
ACCESSION	AX480724				
VERSION	AX480724.1	GI:22217466			
KEYWORDS	human.				
SOURCE	Homo sapiens				
ORGANISM	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Homindae; Homo.				
REFERENCE	Machina,R.A., Recipdon,H., Pluta,J., Ghosh,M.G., Sun,Y. and Liu,C. Compositions and methods relating to colon specific genes and proteins				
AUTHORS	Patent: WO 0242460-A 14 30-MAY-2002;				
JOURNAL	Diadexus, Inc. (US)				
FEATURES	location/Qualifiers				
source	1..470				
	/organism="Homo sapiens"				
	/db_xref="taxon:9606"				
BASE COUNT	98 a 122 c 147 g 103 t				
ORIGIN					

	Query Match	Best Local Similarity	Score	DB	Length
	Matches	466;	Conservative	0;	Pred. No. 8e-111;
					Fmi. 1; Indels 0; Gaps 0;
QY	567	CAGCAGCTGGGCGTTCCTGTGGAGGGAGTGACTTGCACTGGCAGACACATGCAATGTCACCTG	626		
Db	2	CAGGTGCTGGGCTTCCCTGTGAGGGAGTAGCTTGCACTGGCAGACACATGCAATGTCACCTG	61		
QY	627	GGAAACCCCTGGAGCAAAAGCTACATCCCAAGACAGACAGATGATGATCCAGAGCAAAACGGC	686		
Db	62	GGAAACCCCTGGAGCAAAAGCTACATCCCAAGACAGACAGATGATGATCCAGAGCAAAACGGC	121		
QY	687	AATATATGCCAATGTTAAATGTGATGATTACACAGCTAGCTATGAGGATGCTGCTCTTA	746		
Db	122	AATATATGCCAATGTTAAATGTGATGATTACACAGCTAGCTATGAGGATGCTGCTCTTA	181		
QY	747	GTCCAGGAAATATGGGGGTATGATGCTGCTTCCAACTCTGTGGCTGTAAAGCAAGCTCAG	806		
Db	182	GTCCAGGAAATATGGGGGTATGATGCTGCTTCCAACTCTGTGGCTGTAAAGCAAGCTCAG	241		
QY	807	GCTAGTCTCCCACTGGGGGCTGTGCCCTCCCTGGGAGGGTTCCGGGGGACAGCCCATC	866		
Db	242	GCTAGTCTCCCACTGGGGGCTGTGCCCTCCCTGGGAGGGTTCCGGGGGACAGCCCATC	301		

OY	867	ACTGTCCTCAATAGGTGTGAGATGTATGCTAAAGCCCTGCTGTGCTGTGCTGCACATGCCA	926
Db	302	ACTGTCCTCAATAGGTGTGAGATGTATGCTAAAGCCCTGCTGTGCTGTGCTGCACATGCCA	361
OY	927	CAGCAGCGGTGGGGCTCGTGGGGACATCCATCGTGAGTGTCTCTCAGCCTTAGT	986
Db	362	CAGCAGCGGTGGGGCTCGTGGGGACATCCATCGTGAGTGTCTCTCAGCCTTAGT	421
OY	987	CTGACAGAGACTTGGCGGGGATGCTCCAGGATGTGGGTGATTTCTGT	1035
Db	422	CTGACAGAGACTTGGCGGGGATGCTCCAGGATGTGGGTGATTTCTGT	470

LOCUS	AX408491/c			
DEFINITION	Sequence 1138 from Patent WO0229103.	427 bp	DNA	linear
ACCESSION	AX408491			
VERSION	AX408491.1	GI:24441196		
KEYWORDS				
SOURCE	human.			
ORGANISM	Homo sapiens			
REFERENCE	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.			
AUTHORS	Alvares, C., Horne, D., Peres-da-Silva, S. and Vockley, J.G.			
TITLE	Gene expression profiles in liver cancer			
JOURNAL	Patent: WO 0229103-A 1138 11-APR-2002;			
	GENE LOGIC INC (US)			

FEATURES	Location/Qualifiers
source	1..427
	/organism="Homo sapiens"
	/db_xref="taxon:9606"
	/note="EMBL/GenBank Accession No. AA451877"
BASE COUNT	107 a 126 c 92 g 102 t
ORIGIN	

Query Match	30.1%	Score 420.8	DB 6	Length 427
Best Local Similarity	99.5%	Pred: No. 2e-98		
Matches 422, Conservative	0	Mismatches 2	Indels 0	Gaps 0

Accession	Sequence	Position
Oy	GTGGGGGCGTGCCTGGGGGCAATTCATCGTGAGATGTTCCCTCAGTTAGTGTGACAG	995
Db	GTGGGGGCGTGCCTGGGGCAATTCATCGTGAGATGTTCTCTCAGCTTAGTGTGACAG	368
Oy	AGACTTGGCGGGGAGTGTCTCCAGAGATGTGGGATTCGTGACCTGGGAGGCTAATCTG	1055
Db	AGACTTGGCGGGAGATGTCTCCAGAGATGTGGGATTCGTGACCTGGGAGGCTAATCTCTG	308

QY 1056 ACCTCCGACAGGGGACACTCCGACGGCCAGGCCAGGGGTCAAGGGCAGAGTGCACACT 1115

1116 CAGCATGAGCCACAGACTGGGGTCAGGGAGCAGCTGTGTTGAGCCAGACACTGGGGCGS 1175

**Z**

**ZC** GAGGTCGGCCCGGCCCTTCTGCCTCATATGAACCTCAAAAGCACCCAAA 123

**ZY** CAGCACAAGACCATTGATCCCCAAGTGAGTTGTACTTAAGGACGACCCGCGC 188

**Dd** 187 GGGTGGACCGGSGCCTTTCGTGCCTCATTTTCCTTCAATGAAGCCCTCAAAAGCAGCCAAA 128

Db 127 ACCAGGCTTCCCCCTCTCGAGTTGAATATCCAGAACTTTTGTACTTCTGTGT 68

67 TAAATGTTTAATTTGTAATAAATAAATTAAGTAAATAAATGATGTTTCACAG 8  
 1236 AAAATGGTATATTTGTAATAAATAAATAAGTAAATAAATGATGTTTCACAG 1355  
 1716 AAAATGGTATATTTGTAATAAATAAATAAGTAAATAAATGATGTTTCACAG 1835

QY	1356	CMAA	1359
MB	7	CHP	4

RESULT 8					
AC122406/c					
LOCUS	AC122406	176771 bp	DNA	linear	HTG 20-JUN-2002
DEFINITION	Mus musculus chromosome UNK clone R224-126114, WORKING DRAFT				

ACCESSION	VERSION	KEYWORDS	SOURCE	ORGANISM
AC122406	GI:21490547	HTG; HTGS_PHASE1; HTGS_DRAFT; HTGS_FULLTOP.	house mouse.	Mus musculus.

REFERENCE  
AUTHORS  
TITLE  
JOURNAL  
REFERENCE  
AUTHORS  
TITLE  
JOURNAL

1 (bases 1 to 176771)  
McPherson, J.D. and Waterston, R.H.  
The sequence of *Mus musculus* clone  
Unpublished  
2 (bases 1 to 176771)  
McPherson, J.D. and Waterston, R.H.  
Direct Submission  
Submitted (22-MAY-2002) Genome Sequencing Center, 4444 Forest Park

REFERENCE	3 (bases 1 to 176771)
AUTHORS	McPherson, J. D. and Waterston, R. H.
TITLE	Direct Submission
JOURNAL	Submitted (20-JUN-2002) Genome Sequencing Center, 4444 Forest Park Parkway, St. Louis, MO 63108, USA
COMMENT	On Jun 20, 2002 this sequence version replaced gi:21105866.

----- Genome Center -----  
 Center: Washington University Genome Sequencing Center  
 Web site: WUGSC  
 Web site: <http://gsnec.wustl.edu/gsc/index.shtml>  
 Contact: [submit@genome.wustl.edu](mailto:submit@genome.wustl.edu)  
 ----- Project Information -----  
 Center project name: M\_BB012614

```
----- Summary Statistics -----
Sequencing vector: M13: 0%
Sequencing vector: plasmid: 100%
Chemistry: Dye-primer ET: 0% of reads
Chemistry: Dye-terminator Big Dye: 100% of reads
Assembly program: Phrap, version 0.990319
Consensus quality: 171598 bases at least Q40
Consensus quality: 172211 bases at least Q30
Consensus quality: 172337 bases at least Q20
Insert size: 186000; agarose-fp
Insert size: 175614; sum-of-contigs
Quality coverage: 11.87 in Q20 bases; agarose-fp
Quality coverage: 11.02 in Q20 bases; sum-of-contigs
```

\* NOTE: This is a 'working draft' sequence. It currently  
\* consists of 12 contigs. The true order of the pieces  
\* is not known and their order in this sequence record is  
\* arbitrary. Gaps between the contigs are represented as  
\* runs of N, but the exact sizes of the gaps are unknown.  
\* This record will be updated with the finished sequence  
\* as soon as it is available and the accession number will  
\* be preserved.

*	1	57:	contig of 57 bp in length
*	58		gap of unknown length
*	158	3050:	contig of 2893 bp in length
*	3051	3150:	gap of unknown length
*	3151	5182:	contig of 2032 bp in length
*	5183		gap of unknown length
*	5283	14170:	contig of 8888 bp in length
*	14171	14270:	gap of unknown length
*	14271		contig of 7845 bp in length
*	22116	22215:	gap of unknown length
*	22216		contig of 10539 bp in length
*	32755		gap of unknown length
*	32855		contig of 13704 bp in length
*	46558		gap of unknown length
*	46559		gap of unknown length



DEFINITION human STS WI-15430, sequence tagged site.  
 ACCESSION G21939  
 VERSION G21939.1 GI:1342265  
 STS: STS sequence; primer; sequence tagged site.  
 KEYWORDS Homo sapiens STS derived from sequences in dbEST and the Unigene collection.  
 SOURCE Homo sapiens  
 ORGANISM Homo sapiens  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.  
 1 (bases 1 to 462)  
 REFERENCE Hudson T.  
 Whitehead Institute/MIT Center for Genome Research; Physically  
 Mapped STS  
 TITLE Whitehead Institute/MIT Center for Genome Research; Physically  
 JOURNAL Unpublished (1995)  
 COMMENT  
 Contact: Thomas Hudson  
 Whitehead Institute/MIT Center for Genome Research  
 Whitehead Institute for Biomedical Research  
 9 Cambridge Center, Cambridge MA 02142 USA  
 Tel: 617 252 1900  
 Fax: 617 252 1902  
 Email: thudson@genome.wi.mit.edu  
 Primer A: AGGAGAAGTTTCTGTGAA  
 Primer B: GCCTCAAGACAGCAAAAC  
 STS size: 150  
 PCR Profile:  
 Presoak:  
 Denaturation:  
 Annealing: 56 degrees C  
 Polymerization:  
 PCR Cycles: 35  
 Thermal Cycler:  
 Protocol:  
 Template: 10 ng  
 Primer: each 5 pM  
 dNTPs: each 4 mM  
 Taq Polymerase: 0.025 units/ul  
 Total Vol: 20 ul  
 Buffer:  
 MgCl2: 1.5 mM  
 KCl: 50 mM  
 Tris-HCl: 10 mM  
 pH: 9.3  
 Derived from dbEST (genbank accession R47986).  
 Location/Qualifiers  
 source  
 1.462  
 /organism="Homo sapiens"  
 /db\_xref="taxon:9606"  
 /map="192.9 CR from top of Chr2 linkage group"  
 STS  
 primer\_bind  
 1..20  
 complement(132..150)  
 primer\_bind  
 107 a 123 c 117 g 108 t 7 others  
 BASE COUNT  
 ORIGIN  
 Query Match 18.2%; Score 254.4; DB 11; Length 462;  
 Best Local Similarity 94.5%; Pred. No. 1.1e-55;  
 Matches 307; Conservative 0; Mismatches 12; Indels 6; Gaps 4;  
 QY 1050 TCTCTACCTCCGAGAGGAGACATCCC--AGGCCAGCCAGGAGGTCAGGGGCAAGAGT 1107  
 DB 325 TCTGACCTTCCTCCGAGAGGAGACATCCCAGGCCAGCCAGGAGGTCAGGGGCAAGAGT 266  
 QY 1108 GCACA-CCTCAGCATGAG--CCAGACTGGGGTCAGGAGAGGAGTGTGTTGAGCC--AG 1163  
 DB 265 GCACACCTCAGCATGAGGCCAGAGCTGGGGTCAGGAGAGGAGTGTGTTGAGCCAGG 206  
 QY 1164 GACCTGGGGGGGGGGGGGGGGGGCTTTCTGCTCATTTGCTTCAATGAAGCCCTC 1223  
 DB 205 ACCCTGGGGGGGGGGGGGGGGGGCTTTCTGCTCATTTGCTTCAATGAAGCCCTC 146

QY 1224 AAGCAGCCAAACAGAGCTTCCCTTCCTGAGTTGAATATCCAGATCTTTTGA 1283  
 DB 145 AAGCAGCCAAACAGAGCTTCCCTTCCTGAGTTGAATATCCAGATCTTTTGA 86  
 QY 1284 CTCTCTGTTGGTTAAATGTTTATTTTGTAAATAATTAATTAATTAATTAAT 1343  
 DB 85 CTCTCTGTTGGTTAAATGTTTATTTTGTAAATAATTAATTAATTAATTAATTAAT 26  
 QY 1344 GATGTTCAAGCAAACTCTCCCT 1368  
 DB 25 GATGTTCAAGCAAACTCTCCCT 1  
 RESULT 10  
 LOCUS BC022157 1586 bp mRNA linear ROD 07-AUG-2002  
 DEFINITION Mus musculus, clone IMAGE:5134400, mRNA, partial cds.  
 ACCESSION BC022157  
 VERSION BC022157.1 GI:18381057  
 KEYWORDS  
 SOURCE house mouse.  
 ORGANISM Mus musculus  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
 1 (bases 1 to 1586)  
 REFERENCE Strausberg, R.  
 Direct Submission  
 Submitted (24-JAN-2002) National Institutes of Health, Mammalian  
 Gene Collection (MGC), Cancer Genomics Office, National Cancer  
 Institute, 31 Center Drive, Room 11A03, Bethesda, MD 20892-2590,  
 USA  
 NIH-MGC project URL: http://mgc.ncl.nih.gov  
 CONTACT: MGC help desk  
 Email: cgabds-remail.nih.gov  
 Tissue Procurement: Jeffrey E. Green, M.D.  
 CDNA Library Preparation: Life Technologies, Inc.  
 DNA Library Arrayed by: The I.M.A.G.E. Consortium (LNL)  
 BC Cancer Agency, Vancouver, BC, Canada  
 info@bcsc.bc.ca  
 Steven Jones, Jennifer Asano, Ian Bosdet, Yaron Butterfield,  
 Susanna Chan, Readman Chiu, Chris Fjell, Erin Garland, Ran Guin,  
 Leticia Hailo, Martin Krzywinski, Reta Kutsche, Oliver Lee, Soo  
 Sen Lee, Victor Ling, Carrie Mathewson, Candice Mcleary, Steven  
 Ness, Pawan Pandon, Anna-Liisa Pirbhu, Parvaneh Saeedi, Jacqueline  
 Schain, Duane Smallue, Michael Smith, Lorraine Spence, Jeff Scott,  
 Michael Thorne, Miranda Tsai, Natasja van den Bosch, Jill Vardy,  
 George Yang, Scott Zuyderduyn, Marco Marra.  
 Clone distribution: MGC clone distribution information can be found  
 through the I.M.A.G.E. Consortium/LNL at: http://image.lnl.gov  
 Series: IRAX Plate: 57 Row: m Column: 21  
 This clone was selected for full length sequencing because it  
 passed the following selection criteria: Hexamer frequency ORF  
 analysis.  
 FEATURES  
 source  
 1.1586  
 /organism="Mus musculus"  
 /db\_xref="taxon:10090"  
 /map="FVB/N"  
 /clone="IMAGE:5134400"  
 /issue\_type="liver, normal, 5 month old male mouse."  
 /clone\_lib="NCI CGAP\_L19"  
 /lab\_host="DH10B"  
 /note="Vector: pCMW-SPORT6"  
 <1..809  
 /codon\_start=3  
 /product="Unknown (protein for IMAGE:5134400)"  
 /protein\_id="AAH22157.1"  
 /db\_xref="GI:18381058"  
 /translation="IPDLTITTELGAPEVMIQLQTSNEDFRICLSIGRLHLHLAH  
 SPLGVTLLDFRPQFVIVNGELKVTLDIDARVETPTCSADCTLEFPARFSPCS

AGCGGNNKXENLYNARYFFETVLLRHSAPSLRPLDSTVNAITGELAWGVETLAC  
 LETTAHLPRSSGYLONSTSSRAEYRIPDSATODYRCWPSYHGGLLSVFNIAEA  
 IDVCSSHQACRAFFVNTQTFWGRKLVFRTGMQNVDPADAKITTVYVAPG"

BASE COUNT 354 a 455 c 441 g 336 t

Query Match 18.0%; Score 251; DB 10; Length 1586;

Best Local Similarity 67.1%; Pred. No. 1e-54; Indels 118; Gaps 16;

Matches 649; Conservative 0; Mismatches 200; Indels 118; Gaps 16;

420 GTTACCTCTACCTCTGCCCCCAGAGTGGAGCTGCTTTTTCAGACTGTAGTAC 479  
 699 GTTACCTCTACCTCTGCCCCCAGAGTGGAGCTGCTTTTTCAGACTGTAGTAC 758  
 480 CAAGTGTCTCTGATCCCAAGCAACATATGTGAAGGCTCTGGCTGATCTTGA 539  
 759 CAAGTGTCTCTGATCCCAAGCAACATATGTGAAGGCTCTGGCTGATCTTGA 817  
 540 GGGCTGGCTGATCCCAAGCAACATATGTGAAGGCTCTGGCTGATCTTGA 599  
 818 GGGCTGGCTGATCCCAAGCAACATATGTGAAGGCTCTGGCTGATCTTGA 860  
 600 TGCATGCGACGACTGATGTCACCTGGGAACCCCTGAGCAAAAGCTAATCCAGAC 659  
 861 TGCATGCGACGACTGATGTCACCTGGGAACCCCTGAGCAAAAGCTAATCCAGAC 906  
 660 AGACAGATGTACCAAGCAACAAAGTGAATATGTTAAATGTGAATTTAC 718  
 907 CAATGATGTACCAAGCAACAAAGTGAATATGTTAAATGTGAATTTAC 966  
 719 AGCTAGTATGGAAGTGTGCTGCTGATGTCAGCAAAATATGAGGATATGCTGCTC 778  
 967 AGCTAGTATGGAAGTGTGCTGCTGATGTCAGCAAAATATGAGGATATGCTGCTC 1005  
 779 CAACCTGTGGGCTGTAGCA-AGCTGAGCTATGCTCCCACTGGGGCTGTGCTC 837  
 1006 TGAAGCTGTAGGCTGTAGCAAGGCTCAGGCTGTGCTC-----TA 1044  
 838 CTTGGGAGCGTTCGTTGGGAGCGCCCAATCACTGTTCAATATGTGAATATGCTGTA 897  
 1045 ACTGGGAGCGTTCGTTGGGAGCGCCCAATCACTGTTCAATATGTGAATATGCTGTA 1102  
 898 AGCCCTGTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 957  
 1103 AAC-----ACTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1152  
 958 CCAATGAGTGTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1017  
 1153 CAGTCTGAGTGTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1211  
 1018 GATGT 1077  
 1212 -GATGT 1270  
 1078 CAGGCGAGCGGCTGAGGCTGAGGCTGAGGCTGAGGCTGAGGCTGAGGCTGAGGCTGAGGCTGAGGCT 1137  
 1271 TAGGCGAGCGGCTGAGGCTGAGGCTGAGGCTGAGGCTGAGGCTGAGGCTGAGGCTGAGGCTGAGGCT 1328  
 1138 CAGGCGAGCGGCTGAGGCTGAGGCTGAGGCTGAGGCTGAGGCTGAGGCTGAGGCTGAGGCTGAGGCT 1197  
 1329 GAGGCGAGCGGCTGAGGCTGAGGCTGAGGCTGAGGCTGAGGCTGAGGCTGAGGCTGAGGCTGAGGCT 1387  
 1198 CCAATGAGTGTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1233  
 1388 CCAATGAGTGTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1447  
 1234 AAACGAGCGGCTTCCCTCTCTGAGATTTGATATCAAAATCTTTTGTGCTTGTG 1293  
 1448 AGGCGAGCGGCTTCCCTCTCTGAGATTTGATATCAAAATCTTTTGTGCTTGTG 1507  
 1294 GTTAAATGTTTATTTTATTTTATTTTATTTTATTTTATTTTATTTTATTTTATTTTATTTTATTTTATTTT 1352  
 1508 ATTAAATGTTTATTTTATTTTATTTTATTTTATTTTATTTTATTTTATTTTATTTTATTTTATTTTATTTTATTTT 1567

QY 1353 CAGCAAA 1359  
 Db 1568 TGACAAA 1574

# RESULT 11

AX056461 1886 bp DNA linear PAT 13-JAN-2001

Sequence 105 from Patent WO0073469.

AX056461.1 GI:12229168

## DEFINITION

Sequence 105 from Patent WO0073469.

## ACCESSION

AX056461

## KEYWORDS

AX056461.1 GI:12229168

## SOURCE

ORGANISM

Murinae gen. sp.

Murinae gen. sp.

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

Mammalia; Eutheria; Rodentia; Sciognathia; Muridae; Murinae.

1 (bases 1 to 1886)

Plowman, G.D., Martinez, R., Whyte, D. and Sudersanam, S.

Protein Kinases

Patent: WO 0073469-A 105 07-DEC-2000;

Sugen, Inc. (US)

Location/Qualifiers

source

1..1886

/organism="Murinae gen. sp."

/db\_xref="taxon:39108"

BASE COUNT 384 a 559 c 553 g 390 t

ORIGIN

Query Match 16.9%; Score 235.8; DB 6; Length 1886;

Best Local Similarity 65.6%; Pred. No. 1e-50;

Matches 644; Conservative 0; Mismatches 197; Indels 140; Gaps 15;

420 GTTACCTCTACCTCTGCCCCCAGAGTGGAGCTGCTTTTTCAGACTGTAGTAC 479  
 990 GTTACCTCTACCTCTGCCCCCAGAGTGGAGCTGCTTTTTCAGACTGTAGTAC 1049  
 480 CAAGTGTCTCTGATCCCAAGCAACATATGTGAAGGCTCTGGCTGATCTTGA 539  
 1050 CAAGTGTCTCTGATCCCAAGCAACATATGTGAAGGCTCTGGCTGATCTTGA 1108  
 540 GGGCTGGCTGATCCCAAGCAACATATGTGAAGGCTCTGGCTGATCTTGA 599  
 1109 GGGCTGGCTGATCCCAAGCAACATATGTGAAGGCTCTGGCTGATCTTGA 1151  
 600 TGCATGCGACGACTGATGTCACCTGGGAACCCCTGAGCAAAAGCTAATCCAGAC 659  
 1152 TGCATGCGACGACTGATGTCACCTGGGAACCCCTGAGCAAAAGCTAATCCAGAC 1197  
 660 AGACAGATGTACCAAGCAACAAAGTGAATATGTTAAATGTGAATTTAC 718  
 1198 CAATGATGTACCAAGCAACAAAGTGAATATGTTAAATGTGAATTTAC 1257  
 719 AGCTAGTATGGAAGTGTGCTGCTGATGTCAGCAAAATATGAGGATATGCTGCTC 778  
 1258 AGCTAGTATGGAAGTGTGCTGCTGATGTCAGCAAAATATGAGGATATGCTGCTC 1296  
 779 CAACCTGTGGGCTGTAGCA-AGCTGAGCTATGCTCCCACTGGGGCTGTGCTC 837  
 1297 TGAAGCTGTAGGCTGTAGCAAGGCTCAGGCTGTGCTC-----TA 1335  
 838 CTTGGGAGCGTTCGTTGGGAGCGCCCAATCACTGTTCAATATGTGAATATGCTGTA 897  
 1336 ACTGGGAGCGTTCGTTGGGAGCGCCCAATCACTGTTCAATATGTGAATATGCTGTA 1393  
 898 AGCCCTGTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 957  
 1394 AAC-----ACTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1443  
 958 CCAATGAGTGTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1017  
 1444 CAGTCTGAGTGTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1501

Oy		1018	CGATGTGGGTCATTCTTGACCTGGGGAGGCTATCTCTCAACCTCCCGA CAGGGGACAATC	1077
Oy		1502	AGATGTGACTGAAC CAGCATCTGGGAAGAAGGCTCTCGGCGCTTCTCC TAGAGGTGCTCC	1561
Oy		1078	CAGGCCAGCCAGGGGTACAGGGGCAGAGTGCACACTCAGCATGAGCCAAGACTGGAGT	11379
Db		1562	TAGCCAGAGCCCCGAGGCCGTGGGGCAGCAGTGTCTGCATTCATATGAGCCMAGACTAGAGT	16212
Oy		1138	CAGGAGCAGCGTGTGTGTTTAGACCAGAACCTGGGGCGGGGGTGGGGCCGGGGCTTTCTG	11977
Db		1622	GGAGGAGCAGATTTGCACTTTGAGCCACAGA-CTGGGGGTGGGGGGAGGGGTGGGGGCTCTCTG	16808
Oy		1139	CCTCATTTGCTTTCATATGAAGCCCTCAAGCACGCC-----	12322
Db		1681	CTCATTTGCTTTCAGTAAGAACCCAGGGAGCAGCCGACAGCCAGGCTCTCCCACTCTGG	17400
Oy		1233	-----AAACACAGGCTTTCCCGCTCTCTCAGTTGAAATATC	12659
Db		1741	AGCCAGAGGCTCTCCCGCTCTCGAGGCGCAGGCTCTCCCGCTCTCGAGTTGGTACC	18000
Oy		1270	CAGAACTCTTTGCTACTCTTCTGTTGTTAAATTTGTTATTTTTGTAATAATTAATAA	13299
Db		1801	CAGAGCTTTTATCTCTCTCGTCAATPAATTTGTTATTTGTAATAAAAAAT-TAA	18559
Oy		1330	TTAGTTAATTAATATGATGTTT	1350
Db		1860	TCAAATTAATTAATATGATGTTT	1880
RESULT 12				
LOCUS	AC099468/c	278726 bp	DNA	linear HTG 12-JUL-2002
DEFINITION	Rattus norvegicus clone CH230-37P5, *** SEQUENCING IN PROGRESS ***,			
ACCESSION	AC099468			
VERSION	AC099468.7 GI:21730074			
KEYWORDS	HTG; HTGS; PHASE1.			
SOURCE	Norway rat.			
ORGANISM	Rattus norvegicus			
	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;			
	Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;			
	Rattus.			
REFERENCE	1 (bases 1 to 278726)			
AUTHORS	Muzny,D.M., Adams,C., Adio-Oduola,B., Ali-osman,F.R., Allen,C.;			
	Alsbrooks,S.L., Amarantunge,H.C., Are,J.R., Ayale,M., Banks,T.,			
	Barbaria,J., Benton,J., Bimage,K.C., Blankenburg,K., Bonnin,D.,			
	Bouck,J., Bowls,S., Brieva,M., Brown,E., Brown,M., Bryant,N.P.,			
	Carroll,T.F., Carter,M., Cavazos,S.R., Currell,K.L., Byrd,N.C.,			
	Chen,G., Chen,R., Chen,Z., Chowdhry,I., Christopoulos,C.,			
	Cleveland,C.D., Cox,C., Coyte,M.D., Dathorne,S.R., David,R.,			
	Davila,K.R., Davis,C., Davy-Carroll,L., Dederich,D.A.,			
	DeLaney,M.R., Delgado,O., Denn,A.B., Ding,Y., Dinh,H.H.,			
	Douthwaite,K.J., Draper,H., Dugan-Kocha,S., Durbin,K.J.,			
	Earnhart,C., Edgar,D., Edwards,C.C., Elhaj,C., Escotto,M.,			
	Falls,T., Ferraguto,D., Flagg,N., Ford,J., Foster,P., Frantz,P.,			
	Gabisl,A., Gao,J., Garcia,A., Garner,T., Garza,N., Gall,R.,			
	Gorelli,J.H., Guevara,W., Gunaratne,P., Hale,S., Hamilton,K.,			
	Harris,C., Harris,K., Hart,M., Haylak,P., Hawes,A., Hernandez,J.,			
	Hernandez,O., Hodgson,A., Hogues,M., Holloway,C., Hollins,B.,			
	Hornel,F., Howard,S., Huber,J., Hulyk,S., Hume,J., Jackson,L.E.,			
	Jacobson,B., Jia,Y., Johnson,R., Jolivet,S., Joudah,S.,			
	Karlsson,E., Kelly,S., Khan,U., King,L., Korvah,J., Kovac,C.,			
	Kratovic,Z., Kureshi,A., Landry,N., Leal,B., Lewis,L.C., Lewis,			
	Li,J., Li,Z., Lichtarge,O., Lieu,C., Liu,J., Liu,W., Louissegh,H.,			
	Lozado,R.U., Lu,X., Lucier,A., Lucier,R., Luna,R., Ma,J.,			
	Mathewswari,M., Mapua,P., Martin,R., Martindale,A., Martinez,E.,			
	Maseley,E., Maxwelline,E., McLeod,M.P., Meador,M., Mei,G., Metzker,M.,			
	Miner,G., Miner,Z., Mitchell,T., Mohabbat,K., Morgan,M., Morris,S.,			
	Moser,M., Neal,D., Newton,J., Newton,N., Nguyen,A., Nguyen,N.,			
	Nguyen,N., Nickerson,E., Nockenkwos,S., Ogun,W., Okunonu,G.,			
	Orangunye,N., Oviedo,R., Pace,A., Payton,B., Peery,J., Perez,L.,			
	Peters,L., Pickens,R., Primus,E., Pu,L.B., Quiles,M., Ren,Y.,			

TITLE.  
JOURNAL  
REFERENCE

AUTHORS  
TITLE  
JOURNAL

REFERENCE  
AUTHORS  
TITLE  
JOURNAL

COMMENT

Ives, M., Rojås, A., Rojshokan, I., Rolfe, M., Ruiz, S., Savery, G., Scherer, S., Scott, G., Shen, H., Shoocharrin, N., Siason, I., Sodergren, E., Sonaike, T., Sparks, A., Stanley, H., Stone, H., Sutton, A., Svatek, A., Taber, P., Tamerisa, A., Tamerisa, K., Tay, H., Tenney, J., Taylor, C., Taylor, T., Terfild, B., Thomas, N., Thomas, S., Uemami, K., Vasquez, L., Vera, V., Villalón, D., Vinson, R., Wang, Q., Wang, S., Ward-Moore, S., Warren, R., Washington, C., Wellington, S., Williams, G., Williamson, A., Wleczyk, R., Wooden, S., Worley, K., Wu, C., Wu, Y., Wu, Y.F., Zhou, J., Zortilla, S., Nelson, D., Weinstein, G., and Gibbs, R.

Unpublished  
Direct Submission  
2 (bases 1 to 278726)  
Worley, K.C.  
Submitted (15-NOV-2001) Human Genome Sequencing Center, Department of Molecular and Human Genetics, Baylor College of Medicine, One Baylor Plaza, Houston, TX 77030, USA  
3 (bases 1 to 278726)  
Worley, K.C.  
Direct Submission  
Submitted (12-JUL-2002) Human Genome Sequencing Center, Department of Molecular and Human Genetics, Baylor College of Medicine, One Baylor Plaza, Houston, TX 77030, USA  
On Jul 11, 2002 this sequence version replaced gi:18874145.

-----  
Genome Center  
Baylor College of Medicine  
Center code: BCM  
Web site: http://www.hgsc.bcm.tmc.edu/  
Contact: hgsc-help@bcm.tmc.edu  
Project information  
Center project name: GIVL  
Center clone name: CH230-37P5  
Summary Statistics  
Sequencing vector: Plasmid;  
Chemistry: Dye-terminator Big Dye; 100k of reads  
Assembly program: Phrap; version 0.990329  
Consensus quality: 209820 bases at least Q40  
Consensus quality: 210653 bases at least Q30  
Consensus quality: 211433 bases at least Q20  
-----

\* NOTE: Estimated insert size may differ from sequence length  
\* (see [http://www.hgsc.bcm.tmc.edu/docs/genbank\\_draft\\_data.html](http://www.hgsc.bcm.tmc.edu/docs/genbank_draft_data.html))  
\* NOTE: This is a "working draft" sequence. It currently  
\* consists of 56 contigs. The true order of the pieces  
\* is not known and their order in this sequence record is  
\* arbitrary. Gaps between the contigs are represented as  
\* runs of N, but the exact sizes of the gaps are unknown.  
\* This record will be updated with the finished sequence  
\* as soon as it is available and the accession number will  
\* be preserved.

1  
1086: contig of 1086 bp in length  
1087 1186: gap of unknown length  
1187 2248: contig of 1062 bp in length  
2249 2348: gap of unknown length  
2349 3500: contig of 1152 bp in length  
3501 3600: gap of unknown length  
3601 4631: contig of 1031 bp in length  
4632 4731: gap of unknown length  
4732 5800: contig of 1069 bp in length  
5801 5900: gap of unknown length  
5901 6902: contig of 1002 bp in length  
6903 7002: gap of unknown length  
7003 8056: contig of 1054 bp in length  
8057 8156: gap of unknown length  
8157 9348: contig of 1192 bp in length  
9349 9448: gap of unknown length  
9449 10517: contig of 1069 bp in length  
10518 10617: gap of unknown length  
10618 11662: contig of 1045 bp in length  
11663 11762: gap of unknown length  
11763 12956: contig of 1194 bp in length  
12957 13056: gap of unknown length



Query Match	Best Local Similarity	Matches	Conservative	Score	DB 2:	Length	278726;
60206	60305:	gap of unknown length					
60306	61616:	contig of 1311 bp in length					
61617	61716:	gap of unknown length					
61717	63181:	contig of 1465 bp in length					
63182	63281:	gap of unknown length					
63282	66815:	contig of 5534 bp in length					
66816	68915:	gap of unknown length					
68916	82280:	contig of 13375 bp in length					
82281	82350:	gap of unknown length					
82391	115109:	contig of 32719 bp in length					
115110	115209:	gap of unknown length					
115210	143234:	contig of 28025 bp in length					
4.1%;	Score 56.6;	DB 2:	Length 278726;				
61.0%;	Pred. No. 0.0026;						
0;	Mismatches 57;	Indels	0;	Gaps	0;		

RESULT 13	AC117061	200207 bp	DNA	linear	HTG 18-JUL-2002
LOCUS	AC117061				
DEFINITION	Rattus norvegicus clone CH230-403121, ***	SEQUENCING IN PROGRESS			
ACCESSION	AC117061				
VERSION	AC117061.2	GI:21746605			
KEYWORDS	HTG; HTGS PHASE1.				
SOURCE	HTG; HTGS PHASE1.				
ORGANISM	Rattus norvegicus				
	Eukaryote; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.				
	1 (bases 1 to 200207)				
	Murphy,D.M., Adams,C., Adio-Oduola,B., Ali-osman,F.R., Allen,C., Alshrooke,S.L., Amarasinge,H.C., Aze,J.R., Ayela,M., Banks,T., Barbata,J., Benton,J., Bimge,K., Blankenburg,K., Bonini,D., Bouck,J., Bowie,S., Brieva,M., Brown,E., Brown,M., Bryant,N.P., Burch,C., Burch,P., Burkett,C., Burrell,K.L., Byrd,N.C., Caron,T.F., Carter,M., Cavazos,S.R., Chacko,J., Chavez,D., Chen,G., Chen,R., Chen,Z., Chowhry,I., Christopoulos,C., Cleveland,C.D., Cox,C., Coyle,M.D., Dathorne,S.R., David,R., Davis,M.L., Davis,C., Davy-Carroll,L., Dederich,D.A., Delaney,K.R., Delgado,O., Denn,A.L., Ding,Y., Dinh,H.H., Douthwaite,K.J., Draper,H., Dugan-Rocha,S., Durbin,K.J., Earhart,C., Edgar,D., Edwards,C.C., Elhaj,C., Escotto,M., Falls,T., Ferraguto,D., Flegel,N., Ford,J., Foster,P., Frantz,P., Galisti,A., Gao,J., Garcia,A., Garner,T., Garza,N., Gill,R., Gottell,J.H., Guevara,M., Gunaratne,P., Hale,S., Hamilton,K., Harris,C., Harris,K., Hart,M., Havlik,P., Haves,A., Hernandez,J., Hernandez,O., Hodgson,A., Hogues,M., Holloway,C., Hollins,B., Homsi,F., Howard,S., Huber,J., Huijck,S., Hume,J., Jackson,L.E., Jacobson,B., Jia,Y., Johnson,R., Jolivet,S., Joudah,S., Karlsson,E., Kelly,S., Khan,U., King,L., Korvah,J., Kovar,C., Kratovic,J., Kureishi,A., Landry,N., Leal,B., Lewis,L.C., Lewis,L., Li,J., Li,Z., Licharge,O., Lieu,C., Liu,J., Liu,W., Loulsegod,H., Lozado,R.J., Lu,X., Lucier,A., Lucier,R., Luna,R., Ma,J., Maheshwari,M., Mapua,P., Martin,R., Mattindale,A., Martinez,E., Massey,E., Mawhinney,E., McLeod,M.P., Meador,M., Mel,G., Metzker,M., Miner,G., Minner,Z., Mitchell,T., Mchabbat,K., Morgan,M., Morris,S., Moser,M., Neal,D., Newton,J., Newton,S., Nguyen,A., Nguyen,N., Nguyen,N., Nickerson,E., Nwokkenko,S., Ogum,H., Okunomi,G., Orangunye,N., Oyiedo,R., Pace,A., Payton,B., Peery,J., Perez,L.,				



Peters, L., Pickens, R., Primus, E., Pu, L. L., Quiles, M., Ren, Y.,  
 Rivers, M., Rojase, A., Rojupokan, I., Rolle, M., Ruiz, S., Savery, G.,  
 Scherer, S., Scott, G., Shen, H., Shoshitari, N., Sisson, I.,  
 Sodegren, E., Sonaike, T., Sparks, A., Stanley, H., Stone, H.,  
 Sutton, A., Svatek, A., Tabor, P., Tamerisa, A., Tamerisa, K., Tang, H.,  
 Tansley, J., Taylor, C., Taylor, T., Telford, B., Thomas, N., Thomas, S.,  
 Umami, K., Vasquez, L., Vera, V., Villalon, D., Vinson, R., Wang, Q.,  
 Wang, S., Ward-Moore, S., Warren, R., Washington, C., Wallington, S.,  
 Williams, G., Williamson, A., Wleczky, R., Wooden, S., Worley, K.,  
 Wu, C., Wu, Y., Wu, Y. F., Zhou, J., Zorrilla, S., Nelson, D.,  
 Weinstein, G., and Gibbs, R.

Unpublished  
 Direct Submission  
 2 (bases 1 to 200207)  
 Worley, K.C.

Submitted (06-APR-2002) Human Genome Sequencing Center, Department  
 of Molecular and Human Genetics, Baylor College of Medicine, One  
 Baylor Plaza, Houston, TX 77030, USA  
 3 (bases 1 to 200207)

Worley, K.C.  
 Direct Submission  
 Submitted (18-JUL-2002) Human Genome Sequencing Center, Department  
 of Molecular and Human Genetics, Baylor College of Medicine, One  
 Baylor Plaza, Houston, TX 77030, USA  
 On Jul 14, 2002 this sequence version replaced gi:20066112.

Genome Center  
 Center: Baylor College of Medicine  
 Center code: BCM  
 Web site: <http://www.hgsc.bcm.tmc.edu/>  
 Contact: [hgsc-help@bcm.tmc.edu](mailto:hgsc-help@bcm.tmc.edu)  
 Project Information  
 Center project name: CH230-403121  
 Center clone name: CH230-403121

Summary Statistics  
 Sequencing vector: Plasmid;  
 Chemistry: Dye-terminator Big Dye 100% of reads  
 Assembly program: Phrap; version 0.990329  
 Consensus quality: 66669 bases at least Q40  
 Consensus quality: 70534 bases at least Q30  
 Consensus quality: 73881 bases at least Q20

NOTE: Estimated insert size may differ from sequence length  
 (see [http://www.hgsc.bcm.tmc.edu/docs/genbank\\_draft\\_data.html](http://www.hgsc.bcm.tmc.edu/docs/genbank_draft_data.html)).  
 NOTE: This is a 'working draft' sequence. It currently  
 consists of 112 contigs. The true order of the pieces  
 is not known and their order in this sequence record is  
 arbitrary. Gaps between the contigs are represented as  
 runs of N, but the exact sizes of the gaps are unknown.  
 This record will be updated with the finished sequence  
 as soon as it is available and the accession number will  
 be preserved.

1307: contig of 1307 bp in length  
 1308 1407: gap of unknown length  
 1408 2750: contig of 1343 bp in length  
 2751 2850: gap of unknown length  
 2851 4107: contig of 1257 bp in length  
 4108 4207: gap of unknown length  
 4207 5462: contig of 1255 bp in length  
 5462 5562: gap of unknown length  
 5562 6941: contig of 1379 bp in length  
 6941 7041: gap of unknown length  
 7041 8325: contig of 1284 bp in length  
 8325 8425: gap of unknown length  
 8425 9428: contig of 1003 bp in length  
 9428 9528: gap of unknown length  
 9528 10838: contig of 1310 bp in length  
 10838 12015: contig of 1077 bp in length  
 12015 12115: gap of unknown length  
 12115 13380: contig of 1265 bp in length  
 13380 13480: gap of unknown length  
 13480 14483: contig of 1003 bp in length

14484 14583: gap of unknown length  
 14583 15814: contig of 1231 bp in length  
 15814 15914: gap of unknown length  
 15914 16914: contig of 1000 bp in length  
 16914 17014: gap of unknown length  
 17014 18252: contig of 1238 bp in length  
 18252 18352: gap of unknown length  
 18352 19639: contig of 1287 bp in length  
 19639 19739: gap of unknown length  
 19739 21007: contig of 1268 bp in length  
 21007 21107: gap of unknown length  
 21107 22132: contig of 1025 bp in length  
 22132 22232: gap of unknown length  
 22232 23307: contig of 1075 bp in length  
 23307 23407: gap of unknown length  
 23407 24549: contig of 1142 bp in length  
 24549 24649: gap of unknown length  
 24649 26033: contig of 1384 bp in length  
 26033 26133: gap of unknown length  
 26133 27327: contig of 1194 bp in length  
 27327 27427: gap of unknown length  
 27427 28447: contig of 1020 bp in length  
 28447 28547: gap of unknown length  
 28547 30342: contig of 1795 bp in length  
 30342 30442: gap of unknown length  
 30442 31486: contig of 1044 bp in length  
 31486 31586: gap of unknown length  
 31586 32588: contig of 1272 bp in length  
 32588 32958: gap of unknown length  
 32958 33977: contig of 1019 bp in length  
 33977 34077: gap of unknown length  
 34077 35589: contig of 1512 bp in length  
 35589 37047: gap of unknown length  
 37047 37147: gap of unknown length  
 37147 38713: contig of 1566 bp in length  
 38713 38813: gap of unknown length  
 38813 39996: contig of 1183 bp in length  
 39996 40096: gap of unknown length  
 40096 41126: contig of 1030 bp in length  
 41126 41226: gap of unknown length  
 41226 42462: contig of 1236 bp in length  
 42462 42562: gap of unknown length  
 42562 43937: contig of 1375 bp in length  
 43937 44037: gap of unknown length  
 44037 45068: contig of 1031 bp in length  
 45068 45168: gap of unknown length  
 45168 46348: contig of 1180 bp in length  
 46348 46448: gap of unknown length  
 46448 47650: contig of 1202 bp in length  
 47650 47751: gap of unknown length  
 47751 49382: contig of 1632 bp in length  
 49382 49482: gap of unknown length  
 49482 50884: contig of 1402 bp in length  
 50884 50984: gap of unknown length  
 50984 52017: contig of 1033 bp in length  
 52017 52117: gap of unknown length  
 52117 53401: contig of 1284 bp in length  
 53401 53501: gap of unknown length  
 53501 54545: contig of 1044 bp in length  
 54545 54645: gap of unknown length  
 54645 55789: contig of 1144 bp in length  
 55789 55889: gap of unknown length  
 55889 56979: contig of 1090 bp in length  
 56979 58407: contig of 1328 bp in length  
 58407 58507: gap of unknown length  
 58507 59633: contig of 1126 bp in length  
 59633 59733: gap of unknown length  
 59733 60812: contig of 1079 bp in length  
 60812 60912: gap of unknown length  
 60912 61915: contig of 1003 bp in length  
 61915 62015: gap of unknown length

```
1. 133501
/organism="Dictyostelium discoideum"
/strain="AX4"
/db_xref="taxon:44689"
/chromosome="2"
map="1512060-164559"
```

AUTHORS  
TITLE  
JOURNAL  
REFERENCE  
AUTHORS  
TITLE  
JOURNAL  
COMMENT

Worley, K.C.  
Direct Submission  
Submitted (21-MAY-2002) Human Genome Sequencing Center, Department  
of Molecular and Human Genetics, Baylor College of Medicine, One  
Baylor Plaza, Houston, TX 77030, USA  
3 (bases 1 to 158007)  
Worley, K.C.  
Direct Submission  
Submitted (23-JUL-2002) Human Genome Sequencing Center, Department  
of Molecular and Human Genetics, Baylor College of Medicine, One  
Baylor Plaza, Houston, TX 77030, USA  
On Jul 18, 2002 this sequence version replaced gi:21039560.

----- Genome Center  
Center: Baylor College of Medicine  
Center code: BCM  
Web site: <http://www.hgsc.bcm.tmc.edu/>  
Contact: [hgsc-help@bcm.tmc.edu](mailto:hgsc-help@bcm.tmc.edu)  
----- Project Information  
Center project name: GWC2  
Center clone name: CH230-435D23  
----- Summary Statistics  
Sequencing vector: Plasmid  
Chemistry: Dye-terminator Big Dye 100% of reads  
Assembly program: Phrap, version 0.950329  
Consensus quality: 101786 bases at least Q40  
Consensus quality: 105464 bases at least Q30  
Consensus quality: 108013 bases at least Q20

-----  
\* NOTE: Estimated insert size may differ from sequence length  
\* (see [http://www.hgsc.bcm.tmc.edu/docs/Genbank\\_draft\\_data.html](http://www.hgsc.bcm.tmc.edu/docs/Genbank_draft_data.html)).  
\* NOTE: This is a 'working draft' sequence. It currently  
\* consists of 55 contigs. The true order of the pieces  
\* is not known and their order in this sequence record is  
\* arbitrary. Gaps between the contigs are represented as  
\* runs of N, but the exact sizes of the gaps are unknown.  
\* This record will be updated with the finished sequence  
\* as soon as it is available and the accession number will  
\* be preserved.

1 1143: contig of 1143 bp in length  
1144 1243: gap of unknown length  
1244 2605: contig of 1362 bp in length  
2606 2705: gap of unknown length  
2706 3983: contig of 1278 bp in length  
3984 4084: gap of unknown length  
4084 5429: contig of 1346 bp in length  
5430 5529: gap of unknown length  
5530 6913: contig of 1384 bp in length  
6914 7013: gap of unknown length  
7014 8256: contig of 1243 bp in length  
8257 8357: gap of unknown length  
8357 9498: contig of 1142 bp in length  
9499 9598: gap of unknown length  
9599 10637: contig of 1039 bp in length  
10638 10737: gap of unknown length  
10738 11892: contig of 1155 bp in length  
11893 11992: gap of unknown length  
11993 13170: contig of 1178 bp in length  
13171 13270: gap of unknown length  
13271 14303: contig of 1033 bp in length  
14304 14403: gap of unknown length  
14404 15701: contig of 1298 bp in length  
15702 15801: gap of unknown length  
15802 16901: contig of 1100 bp in length  
16902 17001: gap of unknown length  
17002 18052: contig of 1051 bp in length  
18053 18152: gap of unknown length  
18153 19776: contig of 1624 bp in length  
19777 19876: gap of unknown length  
19877 21143: contig of 1267 bp in length  
21144 21243: gap of unknown length  
21244 22441: contig of 1198 bp in length  
22442 22541: gap of unknown length  
22542 24034: contig of 1493 bp in length

24035 24134: gap of unknown length  
24135 25621: contig of 1487 bp in length  
25622 25721: gap of unknown length  
25722 27224: contig of 1503 bp in length  
27225 27324: gap of unknown length  
27325 29327: gap of unknown length  
29328 30598: contig of 1271 bp in length  
30599 30698: gap of unknown length  
30699 32237: contig of 1539 bp in length  
32238 32337: gap of unknown length  
32338 33956: contig of 1619 bp in length  
33957 34056: gap of unknown length  
34057 35436: contig of 1380 bp in length  
35437 35536: gap of unknown length  
35537 37417: contig of 1881 bp in length  
37418 37517: gap of unknown length  
37518 39147: contig of 1630 bp in length  
39148 39247: gap of unknown length  
39248 40937: contig of 1690 bp in length  
40938 41037: gap of unknown length  
41038 43441: contig of 2404 bp in length  
43442 43541: gap of unknown length  
43542 45283: contig of 1742 bp in length  
45284 45383: gap of unknown length  
45384 47906: contig of 2523 bp in length  
47907 48006: gap of unknown length  
48007 50783: contig of 2777 bp in length  
50784 50883: gap of unknown length  
50884 53210: contig of 2327 bp in length  
53211 53310: gap of unknown length  
53311 56392: contig of 3082 bp in length  
56393 56492: gap of unknown length  
56493 59673: contig of 3181 bp in length  
59674 59773: gap of unknown length  
59774 62824: contig of 3051 bp in length  
62825 62924: gap of unknown length  
62925 66675: contig of 3751 bp in length  
66676 66775: gap of unknown length  
66776 70074: contig of 3299 bp in length  
70075 70174: gap of unknown length  
70175 72804: contig of 2630 bp in length  
72805 72904: gap of unknown length  
72905 76454: contig of 3550 bp in length  
76455 76554: gap of unknown length  
76555 80449: contig of 3895 bp in length  
80450 80549: gap of unknown length  
80550 83438: contig of 2889 bp in length  
83439 83539: gap of unknown length  
83540 88068: contig of 4530 bp in length  
88069 88168: gap of unknown length  
88169 93861: contig of 5693 bp in length  
93862 93961: gap of unknown length  
93962 98834: contig of 4872 bp in length  
98835 98934: gap of unknown length  
98935 102910: contig of 3976 bp in length  
102911 103010: gap of unknown length  
103011 106942: contig of 3932 bp in length  
106943 107042: gap of unknown length  
107043 110510: contig of 3468 bp in length  
110511 110610: gap of unknown length  
110611 114467: contig of 3857 bp in length  
114468 114567: gap of unknown length  
114568 118093: contig of 3526 bp in length  
118094 118193: gap of unknown length  
118194 125078: contig of 6885 bp in length  
125079 125178: gap of unknown length  
125180 130442: contig of 5264 bp in length  
130443 130542: gap of unknown length  
130543 136904: contig of 6362 bp in length  
136905 137004: gap of unknown length  
137005 143575: contig of 6571 bp in length

Query Match 3.8; Score 52.8; DB 2; Length 15807;  
Best Local Similarity 51.3; Pred. No. 0.024;  
Matches 117; Conservative 0; Mismatches 111; Indels 0; Gaps 0;

QY	1169	GGGCGGGGGTGGGCGCGGCGCTTCTGCTCATTTGCTTCATGAAAGCCTCAAGC	1228
Db	2281	GGGCGGGGGGGGGGGGGGTTTTCCTCCNCCCCCCCCCCCCCGGGGGGGGG	2222
QY	1229	AGCCAAACGAGGCTTTCCTTCCTCGAGTTGAATATCGAGATCTTTGACTTCT	1288
Db	2221	AAAAANCCCCCCCCCTTCTTTTNTNAAAAAATTTTGTGATTA	2162
QY	1289	TGTTGTTAATTTGTTAATTTTGTAAAAATTAATTAATTAATTAATTAATGATGT	1348
Db	2161	AAGAAAGGAAATGTTTTTTTTTAAAAAAGGCGGGGGGTTT	2102
QY	1349	TTCACAGCAACTCTCCCTAAAAAAGGCGGT	1396
Db	2101	TTTTTTTTTCCCCCCCCCAAAAAAAGAGGTGT	2054

Search completed: June 29, 2003, 04:55:58  
Job time : 2489 secs

Human prostate ex
DNA encoding novel
Human prostate exp
Human prostate exp
Human prostate exp
Human angiotensin II
Human cancer adenom
Human immune syst
Human prostate exp
Human polynucleoti
Mouse fatty acid e
Human prostate exp
Human prostate exp
Human prostate exp
Human prostate exp
Human prostate exp
Human breast cancer
Human prostate exp
Human secreted pro
Human prostate exp
Chemically treated
Human polynucleoti
Human cervical can
Human prostate exp
Human prostate exp
Human prostate exp
Human prostate exp

PR 17-MAR-2000; 2000US-0190076-  
PR 18-APR-2000; 2000US-0198113-  
PR 19-MAY-2000; 2000US-0205515-  
PR 07-JUN-2000; 2000US-0209467-  
PR 28-JUN-2000; 2000US-0214886-  
PR 30-JUN-2000; 2000US-0215135-  
PR 07-JUL-2000; 2000US-0216647-  
PR 07-JUL-2000; 2000US-0216880-  
PR 11-JUL-2000; 2000US-0217487-  
PR 11-JUL-2000; 2000US-0217496-  
PR 14-JUL-2000; 2000US-0218290-  
PR 26-JUL-2000; 2000US-0220963-  
PR 14-AUG-2000; 2000US-0224518-  
PR 14-AUG-2000; 2000US-0224519-  
PR 14-AUG-2000; 2000US-0225213-  
PR 14-AUG-2000; 2000US-0225214-  
PR 14-AUG-2000; 2000US-0225266-  
PR 14-AUG-2000; 2000US-0225267-  
PR 14-AUG-2000; 2000US-0225268-  
PR 14-AUG-2000; 2000US-0225270-  
PR 14-AUG-2000; 2000US-0225447-  
PR 14-AUG-2000; 2000US-0225757-  
PR 14-AUG-2000; 2000US-0225757-  
PR 14-AUG-2000; 2000US-0225759-  
PR 14-AUG-2000; 2000US-0225759-  
PR 18-AUG-2000; 2000US-0226279-  
PR 22-AUG-2000; 2000US-0226681-  
PR 22-AUG-2000; 2000US-0226868-  
PR 22-AUG-2000; 2000US-0227182-  
PR 23-AUG-2000; 2000US-0227009-  
PR 30-AUG-2000; 2000US-0228924-  
PR 01-SEP-2000; 2000US-0229287-  
PR 01-SEP-2000; 2000US-0229343-  
PR 01-SEP-2000; 2000US-0229344-  
PR 01-SEP-2000; 2000US-0229345-  
PR 05-SEP-2000; 2000US-0229509-  
PR 05-SEP-2000; 2000US-0229513-  
PR 06-SEP-2000; 2000US-0230437-  
PR 08-SEP-2000; 2000US-0230438-  
PR 08-SEP-2000; 2000US-0231242-  
PR 08-SEP-2000; 2000US-0231243-  
PR 08-SEP-2000; 2000US-0231244-  
PR 08-SEP-2000; 2000US-0231413-  
PR 08-SEP-2000; 2000US-0231414-  
PR 08-SEP-2000; 2000US-0232080-  
PR 08-SEP-2000; 2000US-0232081-  
PR 12-SEP-2000; 2000US-0231968-  
PR 14-SEP-2000; 2000US-0232357-  
PR 14-SEP-2000; 2000US-0232398-  
PR 14-SEP-2000; 2000US-0232399-  
PR 14-SEP-2000; 2000US-0232400-  
PR 14-SEP-2000; 2000US-0232401-  
PR 14-SEP-2000; 2000US-0233063-  
PR 14-SEP-2000; 2000US-0233064-  
PR 14-SEP-2000; 2000US-0233065-  
PR 21-SEP-2000; 2000US-0234223-  
PR 21-SEP-2000; 2000US-0234274-  
PR 25-SEP-2000; 2000US-0234997-  
PR 25-SEP-2000; 2000US-0234998-  
PR 26-SEP-2000; 2000US-0235484-  
PR 27-SEP-2000; 2000US-0235634-  
PR 27-SEP-2000; 2000US-0235636-  
PR 29-SEP-2000; 2000US-0236327-  
PR 29-SEP-2000; 2000US-0236327-  
PR 29-SEP-2000; 2000US-0236367-  
PR 29-SEP-2000; 2000US-0236368-  
PR 29-SEP-2000; 2000US-0236369-  
PR 29-SEP-2000; 2000US-0236370-  
PR 02-OCT-2000; 2000US-0236802-  
PR 02-OCT-2000; 2000US-0237037-  
PR 02-OCT-2000; 2000US-0237038-  
PR 02-OCT-2000; 2000US-0237039-  
PR 02-OCT-2000; 2000US-0237040-  
PR 13-OCT-2000; 2000US-0239935-

PR 13-OCT-2000; 2000US-0239937-  
PR 20-OCT-2000; 2000US-0240960-  
PR 20-OCT-2000; 2000US-0241221-  
PR 20-OCT-2000; 2000US-0241785-  
PR 20-OCT-2000; 2000US-0241786-  
PR 20-OCT-2000; 2000US-0241787-  
PR 20-OCT-2000; 2000US-0241808-  
PR 20-OCT-2000; 2000US-0241809-  
PR 20-OCT-2000; 2000US-0241826-  
PR 01-NOV-2000; 2000US-0244617-  
PR 08-NOV-2000; 2000US-0246474-  
PR 08-NOV-2000; 2000US-0246475-  
PR 08-NOV-2000; 2000US-0246476-  
PR 08-NOV-2000; 2000US-0246477-  
PR 08-NOV-2000; 2000US-0246478-  
PR 08-NOV-2000; 2000US-0246523-  
PR 08-NOV-2000; 2000US-0246524-  
PR 08-NOV-2000; 2000US-0246525-  
PR 08-NOV-2000; 2000US-0246526-  
PR 08-NOV-2000; 2000US-0246527-  
PR 08-NOV-2000; 2000US-0246528-  
PR 08-NOV-2000; 2000US-0246532-  
PR 08-NOV-2000; 2000US-0246532-  
PR 08-NOV-2000; 2000US-0246609-  
PR 08-NOV-2000; 2000US-0246610-  
PR 08-NOV-2000; 2000US-0246611-  
PR 17-NOV-2000; 2000US-0246613-  
PR 17-NOV-2000; 2000US-0249207-  
PR 17-NOV-2000; 2000US-0249208-  
PR 17-NOV-2000; 2000US-0249309-  
PR 17-NOV-2000; 2000US-0249310-  
PR 17-NOV-2000; 2000US-0249311-  
PR 17-NOV-2000; 2000US-0249312-  
PR 17-NOV-2000; 2000US-0249313-  
PR 17-NOV-2000; 2000US-0249314-  
PR 17-NOV-2000; 2000US-0249315-  
PR 17-NOV-2000; 2000US-0249316-  
PR 17-NOV-2000; 2000US-0249317-  
PR 17-NOV-2000; 2000US-0249318-  
PR 17-NOV-2000; 2000US-0249319-  
PR 17-NOV-2000; 2000US-0249324-  
PR 17-NOV-2000; 2000US-0249325-  
PR 17-NOV-2000; 2000US-0249326-  
PR 17-NOV-2000; 2000US-0249327-  
PR 17-NOV-2000; 2000US-0249329-  
PR 17-NOV-2000; 2000US-0249330-  
PR 01-DEC-2000; 2000US-0250160-  
PR 05-DEC-2000; 2000US-0251030-  
PR 05-DEC-2000; 2000US-0251988-  
PR 05-DEC-2000; 2000US-0256719-  
PR 06-DEC-2000; 2000US-0251479-  
PR 08-DEC-2000; 2000US-0251856-  
PR 08-DEC-2000; 2000US-0251868-  
PR 08-DEC-2000; 2000US-0251869-  
PR 08-DEC-2000; 2000US-0251989-  
PR 08-DEC-2000; 2000US-0251990-  
PR 11-DEC-2000; 2000US-0256097-  
PR 05-JAN-2001; 2001US-0259678-  
XX  
XX  
PA (HUMA-) HUMAN GENOME SCI INC.  
XX  
XX  
PI Rosen CA, Barash SC, Ruben SM;  
XX  
XX WPI; 2001-465569/50.  
XX  
XX Isolated nucleic acid molecule encoding excretory system antigen is  
XX used in preventing, treating or ameliorating a medical condition -  
PS Example 2; SEQ ID NO 816; 574pp + Sequence Listing; English.  
CC The invention relates to novel excretory system related human  
CC polynucleotides (AA198567-AA199503) and the encoded proteins  
CC (AA99594-AA99913) useful for preventing, treating or ameliorating

CC medical conditions e.g. by protein or gene therapy, especially.  
 CC disorders related to the excretory system. The genes are isolated  
 CC from a range of human tissues disclosed in the specification. The  
 CC nucleic acids, proteins, antibodies and (ant)agonists are useful in  
 CC the diagnosis, treatment and prevention of: (a) cancer, e.g. breast  
 CC and ovarian cancer and other cancers of the adrenal gland, bone, bone  
 CC marrow, breast, gastrointestinal tract, liver, lung, or urogenital;  
 CC (b) immune disorders e.g. Addison's disease, allergies, autoimmune  
 CC haemolytic anaemia, autoimmune thyroiditis, diabetes mellitus, Crohn's  
 CC disease, multiple sclerosis, rheumatoid arthritis and ulcerative  
 CC colitis; (c) cardiovascular disorders such as myocardial ischaemia;  
 CC (d) wound healing; (e) neurological diseases e.g. cerebral anoxia and  
 CC epilepsy; and (f) infectious diseases such as viral, bacterial, fungal  
 CC and parasitic infections.  
 CC Note: The sequence data for this patent did not form part of the  
 CC printed specification, but was obtained in electronic format directly  
 CC from WHO at ftp.wipo.int/pub/published\_pcr\_sequences.  
 CC XX  
 SQ Sequence 9839 BP; 2292 A; 2702 C; 2654 G; 2191 T; 0 other;

Query Match 92.0%; Score 1284.8; DB 22; Length 9839;  
 Best Local Similarity 99.7%; Pred. No. 1.1e-289;  
 Matches 1308; Conservative 0; Mismatches 2; Indels 2; Gaps 2;

QY 59 ACCGAGTACCAAGTATCCAGACAGACCCATCCCGAGAGACTACCGCTGCGCA 118  
 DB 8530 ACAGAGTACCAAGTATCCAGACAGACCCATCCCGAGAGACTACCGCTGCGCA 8589  
 QY 119 TCTTACCACCAAGGAGCTGCTCTTCAAGTATCACTGCTGAGCTGATGTC 178  
 DB 8590 TCTTACCACCAAGGAGCTGCTCTTCAAGTATCACTGCTGAGCTGATGTC 8649  
 QY 179 TGTGAGAGCCATGCGAGTGTGGGGCTTGTGTGACCAACAGACCACTGAGAGGT 238  
 DB 8650 TGTGAGAGCCATGCGAGTGTGGGGCTTGTGTGACCAACAGACCACTGAGAGGT 8709  
 QY 239 GAGCCAGTGGAGAGGCTTCCAGAGAGAGTGGAGAGACTCTCTGAGAGTATGAT 298  
 DB 8710 GAGCCAGTGGAGAGGCTTCCAGAGAGAGTGGAGAGACTCTCTGAGAGTATGAT 8769  
 QY 299 AGTGATCCCCATCGGAGTCAAGAGGGGCTGTGAGTGTGAGAGAGTATACGTGT 358  
 DB 8770 AGTGATCCCCATCGGAGTCAAGAGGGGCTGTGAGTGTGAGAGAGTATACGTGT 8829  
 QY 359 CTTCAAGGAGCAATATTTGAGAGAGTGTGCTTCCCGAGAGAGAAATCCAGCCC 418  
 DB 8830 CTTCAAGGAGCAATATTTGAGAGAGTGTGCTTCCCGAGAGAGAAATCCAGCCC 8889  
 QY 419 TGTTACCTCTACCTGCGCCCGAGTGGAGAGTGTCTTTTCAAGAGTGGATGAG 478  
 DB 8890 TGTTACCTCTACCTGCGCCCGAGTGGAGAGTGTCTTTTCAAGAGTGGATGAG 8949  
 QY 479 CCAAGTGTCTCTGATCCCAAGAACACATATGTGAGAGGCTCTGAGCTATCTG 538  
 DB 8950 CCAAGTGTCTCTGATCCCAAGAACACATATGTGAGAGGCTCTGAGCTATCTG 9008  
 QY 539 AGGGCTCGGCTGACAGCTGATCTCTGAGAGCTGGGCTTCTGTGAGAGGAGTAC 598  
 DB 9009 AGGGCTCGGCTGACAGCTGATCTCTGAGAGCTGGGCTTCTGTGAGAGGAGTAC 9068  
 QY 599 TTGCACTGGAGAGTGCATGTCACTGGAGAGCCCTGAGAGAGAGTATCCAG 658  
 DB 9069 TTGCACTGGAGAGTGCATGTCACTGGAGAGCCCTGAGAGAGAGTATCCAG 9128  
 QY 659 CAGACAGATGTGACAGAGCAAACTGCAATATGCAATATGTTAAATGTGATTACC 718  
 DB 9129 CAGACAGATGTGACAGAGCAAACTGCAATATGCAATATGTTAAATGTGATTACC 9188  
 QY 719 AGCTTGTGATGTGAGAGTGTGCTCTGATGTCAGAGATCTGAGGGGTATGATGCTCTC 778  
 DB 9189 AGCTTGTGATGTGAGAGTGTGCTCTGATGTCAGAGATCTGAGGGGTATGATGCTCTC 9248  
 QY 779 CAACCTGTGGGCTGTGAGAGAGTGTGATGTTCTCCACATGGGGGCTGTGCTCCTCC 838

DB 9249 CAACCTGTGGGCTGTGAGAGAGTGTGATGTTCTCCACATGCGGCTGTGCTCCTCC 9307  
 QY 839 CTGGAGAGGTTCCGTGGAGAGGCTTCACTGTGTTCAATAGTGTAGAGTATGCTAA 898  
 DB 9308 CTGGAGAGGTTCCGTGGAGAGGCTTCACTGTGTTCAATAGTGTAGAGTATGCTAA 9367  
 QY 899 GCCCTGTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 958  
 DB 9368 GCCCTGTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 9427  
 QY 959 CATGTGAGTGTCTCTGAGCTTGTGAGACAGAGACTTGTGGGGGAGTGTCCAG 1018  
 DB 9428 CATGTGAGTGTCTCTGAGCTTGTGAGACAGAGACTTGTGGGGGAGTGTCCAG 9487  
 QY 1019 GATGTGGGATGTTCTTACCTGGAGAGGCTATCTGTGACTCCCGACAGGGAGACTCC 1078  
 DB 9488 GATGTGGGATGTTCTTACCTGGAGAGGCTATCTGTGACTCCCGACAGGGAGACTCC 9547  
 QY 1079 AGGCGAGCCAGGGGTGAGGGGAGAGTGCACACCTGACATGAGCCAAAGACTGGGGT 1138  
 DB 9548 AGGCGAGCCAGGGGTGAGGGGAGAGTGCACACCTGACATGAGCCAAAGACTGGGGT 9607  
 QY 1139 AGGAGCAGTGTGTTGAGCCAGAGCTTGGGGGCGGGGCTGCGGCTTTCTGC 1198  
 DB 9608 AGGAGCAGTGTGTTGAGCCAGAGCTTGGGGGCGGGGCTGCGGCTTTCTGC 9667  
 QY 1199 CTCATTGCTTCAATGAAAGCTCAAGACAGCAAAACAGGCTTTCCCTCTCGA 1258  
 DB 9668 CTCATTGCTTCAATGAAAGCTCAAGACAGCAAAACAGGCTTTCCCTCTCGA 9727  
 QY 1259 GTTGAATATCCAGAACTTTTGTACTCTGTGTTGTTAAATGTTTATTTTGTAA 1318  
 DB 9728 GTTGAATATCCAGAACTTTTGTACTCTGTGTTGTTAAATGTTTATTTTGTAA 9787  
 QY 1319 ATAAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTA 1370  
 DB 9788 ATAAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTA 9839

RESULT 2  
 AA163402  
 ID AA163402 standard; DNA; 9839 BP.  
 XX  
 AC AA163402;  
 XX  
 DT 22-OCT-2001 (first entry)  
 XX  
 DE Human kidney related polynucleotide seq ID NO 717.  
 XX  
 XX Human; kidney antigen; immunosuppressive; antiarthritic; antirheumatic;  
 KW antiproliferative; cytostatic; cardiant; vasodilator; cerebroprotective;  
 KW neurotropic; neuroprotective; antibacterial; virulence; fungicide;  
 KW ophthalmological; antiallergic; hepatocytic; antidiabetic;  
 KW antiinflammatory; antitumor; antineoplastic; anticonvulsant; antiparasitic;  
 KW Gene therapy; cancer; immune disorder; cardiovascular disorder;  
 KW neurological disease; infection; ds.  
 XX  
 OS Homo sapiens.  
 XX  
 PN MO200155323-A2.  
 XX  
 PD 02-AUG-2001.  
 XX  
 XX 17-JAN-2001; 2001WO-US01343.  
 PF  
 XX 31-JAN-2000; 2000US-0179065.  
 PR 04-FEB-2000; 2000US-0180628.  
 PR 24-FEB-2000; 2000US-0184664.  
 PR 02-MAR-2000; 2000US-0186350.  
 PR 16-MAR-2000; 2000US-0189874.  
 PR 17-MAR-2000; 2000US-0190076.  
 PR 18-APR-2000; 2000US-0198123.



PR	19-MAY-2000	2000US-0205657
PR	07-JUN-2000	2000US-0209451
PR	28-JUN-2000	2000US-0214886
PR	30-JUL-2000	2000US-0215153
PR	07-JUL-2000	2000US-0216647
PR	07-JUL-2000	2000US-0216880
PR	11-JUL-2000	2000US-0217486
PR	11-JUL-2000	2000US-0217486
PR	14-JUL-2000	2000US-0218920
PR	26-JUL-2000	2000US-0220964
PR	26-JUL-2000	2000US-0220964
PR	14-AUG-2000	2000US-0224519
PR	14-AUG-2000	2000US-0224519
PR	14-AUG-2000	2000US-0225267
PR	14-AUG-2000	2000US-0225268
PR	14-AUG-2000	2000US-0225270
PR	14-AUG-2000	2000US-0225447
PR	14-AUG-2000	2000US-0225757
PR	14-AUG-2000	2000US-0225758
PR	14-AUG-2000	2000US-0225759
PR	18-AUG-2000	2000US-0226279
PR	22-AUG-2000	2000US-0226681
PR	22-AUG-2000	2000US-0226688
PR	22-AUG-2000	2000US-0227182
PR	23-AUG-2000	2000US-0227909
PR	30-AUG-2000	2000US-0228094
PR	01-SEP-2000	2000US-0229287
PR	01-SEP-2000	2000US-0229343
PR	01-SEP-2000	2000US-0229344
PR	01-SEP-2000	2000US-0229345
PR	05-SEP-2000	2000US-0229509
PR	05-SEP-2000	2000US-0229513
PR	06-SEP-2000	2000US-0230437
PR	06-SEP-2000	2000US-0230438
PR	08-SEP-2000	2000US-0231242
PR	08-SEP-2000	2000US-0231243
PR	08-SEP-2000	2000US-0231444
PR	08-SEP-2000	2000US-0231444
PR	08-SEP-2000	2000US-0231414
PR	08-SEP-2000	2000US-0232080
PR	08-SEP-2000	2000US-0232081
PR	12-SEP-2000	2000US-0233168
PR	12-SEP-2000	2000US-0233297
PR	14-SEP-2000	2000US-0233299
PR	14-SEP-2000	2000US-0233299
PR	14-SEP-2000	2000US-0233400
PR	14-SEP-2000	2000US-0233401
PR	14-SEP-2000	2000US-0233403
PR	14-SEP-2000	2000US-0233665
PR	14-SEP-2000	2000US-0233665
PR	21-SEP-2000	2000US-0234423
PR	21-SEP-2000	2000US-0234423
PR	25-SEP-2000	2000US-0234977
PR	25-SEP-2000	2000US-0234977
PR	26-SEP-2000	2000US-0235498
PR	27-SEP-2000	2000US-0235634
PR	27-SEP-2000	2000US-0235634
PR	29-SEP-2000	2000US-0236327
PR	29-SEP-2000	2000US-0236367
PR	29-SEP-2000	2000US-0236368
PR	29-SEP-2000	2000US-0236369
PR	29-SEP-2000	2000US-0236370
PR	02-OCT-2000	2000US-0236802
PR	02-OCT-2000	2000US-0237037
PR	02-OCT-2000	2000US-0237038
PR	02-OCT-2000	2000US-0237039
PR	02-OCT-2000	2000US-0237040
PR	13-OCT-2000	2000US-0239935
PR	13-OCT-2000	2000US-0239937
PR	20-OCT-2000	2000US-0240960

CC	PR	20-OCT-2000;	2000US-0241321.
CC	PR	20-OCT-2000;	2000US-0241785.
XX	PR	20-OCT-2000;	2000US-0241786.
XX	PR	20-OCT-2000;	2000US-0241787.
DR	PR	20-OCT-2000;	2000US-0241808.
XX	PR	20-OCT-2000;	2000US-0241809.
XX	PR	20-OCT-2000;	2000US-0241826.
XX	PR	01-NOV-2000;	2000US-0244617.
XX	PR	08-NOV-2000;	2000US-0246474.
XX	PR	08-NOV-2000;	2000US-0246475.
PT	PR	08-NOV-2000;	2000US-0246476.
XX	PR	08-NOV-2000;	2000US-0246477.
XX	PR	08-NOV-2000;	2000US-0246478.
XX	PR	08-NOV-2000;	2000US-0246523.
XX	PR	08-NOV-2000;	2000US-0246524.
XX	PR	08-NOV-2000;	2000US-0246525.
XX	PR	08-NOV-2000;	2000US-0246526.
XX	PR	08-NOV-2000;	2000US-0246527.
XX	PR	08-NOV-2000;	2000US-0246532.
XX	PR	08-NOV-2000;	2000US-0246609.
XX	PR	08-NOV-2000;	2000US-0246610.
XX	PR	08-NOV-2000;	2000US-0246611.
XX	PR	08-NOV-2000;	2000US-0246613.
XX	PR	17-NOV-2000;	2000US-0249207.
XX	PR	17-NOV-2000;	2000US-0249208.
XX	PR	17-NOV-2000;	2000US-0249209.
XX	PR	17-NOV-2000;	2000US-0249210.
XX	PR	17-NOV-2000;	2000US-0249211.
XX	PR	17-NOV-2000;	2000US-0249217.
XX	PR	17-NOV-2000;	2000US-0249218.
XX	PR	17-NOV-2000;	2000US-0249244.
XX	PR	17-NOV-2000;	2000US-0249245.
XX	PR	17-NOV-2000;	2000US-0249264.
XX	PR	17-NOV-2000;	2000US-0249265.
XX	PR	17-NOV-2000;	2000US-0249297.
XX	PR	17-NOV-2000;	2000US-0249309.
XX	PR	01-DEC-2000;	2000US-0250160.
XX	PR	01-DEC-2000;	2000US-0250391.
XX	PR	05-DEC-2000;	2000US-0251030.
XX	PR	05-DEC-2000;	2000US-0251988.
XX	PR	05-DEC-2000;	2000US-0256719.
XX	PR	06-DEC-2000;	2000US-0251479.
XX	PR	08-DEC-2000;	2000US-0251856.
XX	PR	08-DEC-2000;	2000US-0251868.
XX	PR	08-DEC-2000;	2000US-0251869.
XX	PR	08-DEC-2000;	2000US-0251989.
XX	PR	08-DEC-2000;	2000US-0251990.
XX	PR	11-DEC-2000;	2000US-0254097.
XX	PR	05-JAN-2001;	2001US-02559678.
XX	PA	(HUMA-).	HUMAN GENOME SCI INC.
XX	PI	Rosen CA,	Batash SC, Ruben SM;
XX	DR	WPI;	2001-488784/53.
XX	PT	New isolated nucleic acids and polypeptides,	useful for diagnosing,
XX	PT	treating and/or preventing human diseases and disorders -	
XX	XX	Disclosure; SEQ ID NO 717; 564pp + Sequence Listing; English.	
CC	CC	The invention relates to novel kidney related polymucleotides	
CC	CC	(AAI62971-AAI6793) and the encoded polypeptides (AAM42417-AAM42691)	
CC	CC	collectively known as kidney antigens and the use of such kidney antigens	
CC	CC	for detecting disorders of the kidney, especially kidney cancer and	
CC	CC	kidney cancer metastases. The polymucleotides and proteins are also	



DR MPI: 2001-476164/51.  
 DR P-PSDB: AAM23513.  
 XX Isolated polypeptide for treatment of diseases, diagnostics, raising  
 PT antibodies and research use -  
 XX  
 XX Claim 1: Page 202-203; 11275bp; English.  
 CC The present invention provides the protein and coding sequences of novel  
 CC proteins from a variety of organisms, including human, dog, cat, horse,  
 CC cow, pig, hamster, monkey, macaque, yeast, bacteria, fruit fly, sea  
 CC urchin and tomato. These were derived from expressed sequence tags (ESTs)  
 CC from the organism of interest. They can be used in diagnostics,  
 CC forensics, gene mapping, identification of mutations, to assess  
 CC biodiversity and for nutritional purposes. The present sequence is a cDNA  
 CC of the invention.  
 XX  
 SQ Sequence 1846 BP; 397 A; 526 C; 538 G; 385 T; 0 other;  
 Query Match 68.2%; Score 952.4; DB 22; Length 1846;  
 Best Local Similarity 84.9%; Pred. No. 2.3e-212;  
 Matches 1171; Conservative 0; Mismatches 1; Indels 208; Gaps 1;  
 QY 1 GGTGCTGACCTGATCCGAGCGGAGCATATCTGCAAGACTCCACGCGAAGACAGCAGTAC 60  
 DB 675 GGTGCTGACCTGATCCGAGCGGAGCATATCTGCAAGACTCCACGCGAAGACAGCAGTAC 734  
 QY 61 CGAGTACCACTGATCCGAGCGGAGCATATCTGCAAGACTCCACGCGAAGACAGCAGTAC 120  
 DB 735 CGAGTACCACTGATCCGAGCGGAGCATATCTGCAAGACTCCACGCGAAGACAGCAGTAC 794  
 QY 121 CTACACACAGGAGGAGCTGCTCTCTTCAAGTTCACCTGCTGAGGCTGTGATGCTG 180  
 DB 795 CTACACACAGGAGGAGCTGCTCTCTTCAAGTTCACCTGCTGAGGCTGTGATGCTG 854  
 QY 181 TGAAGAGCATCCAGTGTGGGCTTTGTGTCAACCAACAGCAACCTGGA CAGGTGA 240  
 DB 855 TGAAGAGCATCCAGTGTGGGCTTTGTGTCAACCAACAGCAACCTGGA CAGGTGA 907  
 QY 241 GCCAGTGGAGAAAGCCCTTCAAGAGGATGGACGAGACTCTCTGAGGTTGATAGTAG 300  
 DB 908 ----- 907  
 QY 301 TGATCCCCCATCGAAGTCAAGAGGAGTGTGATGAGAGAGATATACGTGCT 360  
 DB 908 ----- 907  
 QY 361 TCAAGGAGTCAATTAAGGAGATGCTTGTGCTCCAGAAAGAAATCCAGCCCTG 420  
 DB 908 ----- 907  
 QY 421 TTACCTCTACCTCTGCCCCCAGGTGGAGAGTGTCTTTTTCAGATCTGATGAGCC 480  
 DB 908 ----- 946  
 QY 481 AAGTGTCCCTGATCCCAAGAGCAATATGGAAGGCTCTGCTGCTGCTGATCTGAG 540  
 DB 947 AAGTGTCCCTGATCCCAAGAGCAATATGGAAGGCTCTGCTGCTGCTGATCTGAG 1006  
 QY 541 GGTCTGGCTGACAGCTGATCTATCTCAGAGAGTGGGCTTGTCTGAGAGGAGTACTT 600  
 DB 1007 GGTCTGGCTGACAGCTGATCTATCTCAGAGAGTGGGCTTGTCTGAGAGGAGTACTT 1066  
 QY 601 GCACTGGCAGACCTGCTGCTGCTGAGAAACCTCTGAGCAAAAGTAACTATCCAGACA 660  
 DB 1067 GCACTGGCAGACCTGCTGCTGCTGAGAAACCTCTGAGCAAAAGTAACTATCCAGACA 1126  
 QY 661 GACAGATGTGACAGAGCAAAAGTAAATATGCAAAATGTAAGTAAAGTTTACAG 720  
 DB 1127 GACAGATGTGACAGAGCAAAAGTAAATATGCAAAATGTAAGTAAAGTTTACAG 1186  
 QY 721 CCTAGATGTGAGCTGTGCTCTTATGTCAGAGATCATGGGGTATGACTGCTCTCA 780

DB 1187 CCTAGATGTGAGCTGTGCTCTTATGTCAGAGATCATGGGGTATGACTGCTCTCA 1246  
 QY 781 ACCCTGGAGCTGTAGCAAGCTCAGGCTAGTCTCCCACTGGGGGCTGAGCCCTCCCT 840  
 DB 1247 ACCCTGGAGCTGTAGCAAGCTCAGGCTAGTCTCCCACTGGGGGCTGAGCCCTCCCT 1306  
 QY 841 GGGAGCGTTCCTGGGAGAGCCCATCATCTGTGTCAATATGATGAGAAATGTAAGTAAAG 900  
 DB 1307 GGGAGCGTTCCTGGGAGAGCCCATCATCTGTGTCAATATGATGAGAAATGTAAGTAAAG 1366  
 QY 901 CCTGTGCTGCTGCTGCTGCAATGCAACAGAGCGGCTGCTGCTGCTGCTGCTGCTGCT 960  
 DB 1367 CCTGTGCTGCTGCTGCTGCAATGCAACAGAGCGGCTGCTGCTGCTGCTGCTGCTGCT 1426  
 QY 961 TGTGAGAGTGTCTCTCAGCTTAGAGTGTGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1020  
 DB 1427 TGTGAGAGTGTCTCTCAGCTTAGAGTGTGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1486  
 QY 1021 TGTGAGAGTGTCTCTCAGCTTAGAGTGTGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1080  
 DB 1487 TGTGAGAGTGTCTCTCAGCTTAGAGTGTGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1546  
 QY 1081 GCCAGCCAGAGGAGTCAAGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1140  
 DB 1547 GCCAGCCAGAGGAGTCAAGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1606  
 QY 1141 GAGAGAGAGTGTGAG 1200  
 DB 1607 GAGAGAGAGTGTGAG 1666  
 QY 1201 CATTTGCTTCAATGAAAGCTCAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1260  
 DB 1667 CATTTGCTTCAATGAAAGCTCAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1726  
 QY 1261 TGAATATTCAGAGATCTTTTGTATCTTGTGTTAAATGTTATTTTGTAAATAAT 1320  
 DB 1727 TGAATATTCAGAGATCTTTTGTATCTTGTGTTAAATGTTATTTTGTAAATAAT 1786  
 QY 1321 AAAATTAATTTAGTAAATTAATGATGTTTCAAGAGAACTCTCCCTTAAATAATAATA 1380  
 DB 1787 AAAATTAATTTAGTAAATTAATGATGTTTCAAGAGAACTCTCCCTTAAATAATAATA 1846

RESULT 4  
 ID AAF44725  
 AC AAF44725 standard; cDNA, 2461 BP.  
 XX  
 XX AAF44725;  
 DT 27-MAR-2001 (first entry)  
 XX  
 XX Novel protein kinase cDNA, SEQ ID NO: 106.  
 DE Human; mouse; protein kinase; antiarthritic; antisclerotic; osteopathetic;  
 XX immunosuppressive; candidant; antiinflammatory; antiaslathetic;  
 KW dermatological; antidiabetic; antifertility; gene therapy; vaccine;  
 KW immune disorder; cardiovascular disease; neurodegenerative disease;  
 KW cancer; autoimmune disorder; stroke; inflammatory bowel disease;  
 KW inflammatory pelvic disease; multiple sclerosis; psoriasis; ss.  
 OS Homo sapiens.  
 XX  
 XX MO200073469-A2.  
 PN 07-DEC-2000.  
 PD 26-MAY-2000; 2000WO-US14842.  
 PF 28-MAY-1999; 99US-0136503.  
 PR (SUGEN-) SUGEN INC.  
 XX Plowman GD, Martinez R, Whyte D, Sudersanam S;  
 PA  
 XX

XX WPI; 2001-032161/04.  
 DR P-PSDB; AAB56597.  
 XX  
 PT Nucleic acids encoding kinase polypeptides, useful for diagnosing and  
 PT treating immune-related diseases and disorders, cardiovascular disease,  
 PT neurodegenerative diseases and/or cancers -  
 XX  
 PS Example 1; Fig 2; 310pp; English.  
 XX  
 CC The present sequence encodes a novel protein kinase. The nucleic acids  
 CC and the protein kinases they encode may be used in the treatment and  
 CC diagnosis of diseases associated with inappropriate kinase expression  
 CC such as immune-related diseases and disorders, cardiovascular disease,  
 CC neurodegenerative diseases and/or cancers. The nucleic acids and  
 CC complementary sequences may also be used as DNA probes in diagnostic  
 CC assays. The kinase polypeptides may be used as antigens in the production  
 CC of antibodies of kinase expression and activity. Anti-kinase antibodies  
 CC and kinase antagonists may also be used to down regulate kinase  
 CC expression and activity. Diseases related to kinase expression and  
 CC activity include rheumatoid arthritis, atherosclerosis, autoimmune  
 CC disorders, complications of organ transplantation, myocardial infarction,  
 CC immune disorders, cardiomyopathies, strokes, renal failure,  
 CC oxidative stress related disorders, chronic inflammatory bowel disease,  
 CC chronic inflammatory pelvic disease, multiple sclerosis, asthma,  
 CC osteoarthritis, psoriasis, rhinitis, autoimmunity, diabetes, cancers and  
 CC reproductive disorders.  
 XX  
 SO Sequence 2461 BP; 426 A; 771 C; 811 G; 453 T; 0 other;  
 Query Match 67.3%; Score 940.4; DB 22; Length 2461;  
 Best Local Similarity 84.7%; Pred. No. 1.6e-209;  
 Matches 1159; Conservative 0; Mismatches 1; Indels 208; Gaps 1;  
 QY 1 GGTGCTGCACTGTATCCGAGCGGGAGTATCTGCAAGTCCACGGCAGACGACGTAC 60  
 DB 1302 GGTGCTGCACTGTATCCGAGCGGGAGTATCTGCAAGTCCACGGCAGACGACGTAC 1361  
 QY 61 CGAGTACCACTGTATCCGAGCGGGAGTATCTGCAAGTCCACGGCAGACGACGTAC 120  
 DB 1362 CGAGTACCACTGTATCCGAGCGGGAGTATCTGCAAGTCCACGGCAGACGACGTAC 1421  
 QY 121 CTACACACGAGGAGCTGCTCTTTCAGTGTTCACCTGAGGCTGTGATGCTG 180  
 DB 1422 CTACACACGAGGAGCTGCTCTTTCAGTGTTCACCTGAGGCTGTGATGCTG 1481  
 QY 181 TGAGAGCCATGCCAGTGTGGGCTTTGTGTGTCACCAACCAACCACTGAGAGTGA 240  
 DB 1482 TGAGAGCCATGCCAGTGTGGGCTTTGTGTGTCACCAACCAACCACTGAGAGTGA 1534  
 QY 241 GCCAGTGGAGAGCCCTTCCAGGAGGATGGACGACCTCTGAGAGTTGATAGTAG 300  
 DB 1535 ----- 1534  
 QY 301 TGATCCCCCATCGAAGTCAAGGGGGTGTGAGGTGATGAGAGAGATATAGTGTCT 360  
 DB 1535 ----- 1534  
 QY 361 TCAAGGACATCAATTATGGAGATGCTTTGCTCCAGAAAGAAACATCCAGCCCTG 420  
 DB 1535 ----- 1534  
 QY 421 TTACTCTTCACTCTGCCCCCAGGTGGAGAGTGTCTTTTCAAGACTGATGGAGCC 480  
 DB 1535 ----- 1534  
 QY 481 AAGTGTCCCTGATCCCAACAGACCAATATGTAAGGCTGTGCTGATGATCTGAG 540  
 DB 1574 AAGTGTCCCTGATCCCAACAGACCAATATGTAAGGCTGTGCTGATGATCTGAG 1633  
 QY 541 GGCTCGGCTTACCAAGCTGATATCTGAGAGCTGGGCTTGTGAGAGGAGTGAATT 600  
 DB 1634 GGCTCGGCTTACCAAGCTGATATCTGAGAGCTGGGCTTGTGAGAGGAGTGAATT 1693

QY 601 GCACCTGGACACTGTGATGTCACTTGGAAACCCCTGACAGCAAGCTAACTCCAGACA 660  
 DB 1694 GCACCTGGACACTGTGATGTCACTTGGAAACCCCTGACAGCAAGCTAACTCCAGACA 1753  
 QY 661 GACAGATGGACAGAGCAAAAGTGAATATGCCAATGTTAAATGATGATTTACCG 720  
 DB 1754 GACAGATGGACAGAGCAAAAGTGAATATGCCAATGTTAAATGATGATTTACCG 1813  
 QY 721 CCTAGCTATGGACCTGCTGCTCTTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 780  
 DB 1814 CCTAGCTATGGACCTGCTGCTCTTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1873  
 QY 781 ACCCTGAGGCTGTGTAAGCAAGCTCAGGCTAGTCTCCCACTGGGGCTGTGCCCTCCCT 840  
 DB 1874 ACCCTGAGGCTGTGTAAGCAAGCTCAGGCTAGTCTCCCACTGGGGCTGTGCCCTCCCT 1933  
 QY 841 GGGACGGTTCCTGGGAGCGCCCATCACTGTGTTCATATGTTGAGATGTAGCTAAAGC 900  
 DB 1934 GGGACGGTTCCTGGGAGCGCCCATCACTGTGTTCATATGTTGAGATGTAGCTAAAGC 1993  
 QY 901 CCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 960  
 DB 1994 CCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 2053  
 QY 961 TCGTGAAGTGTCTCTCACTTATGTTGAGTGTGAGACAGAGTGTGAGGAGTGTGAGGAG 1020  
 DB 2054 TCGTGAAGTGTCTCTCACTTATGTTGAGTGTGAGACAGAGTGTGAGGAGTGTGAGGAG 2113  
 QY 1021 TGTGGGTATTTCTGTAACCTGTTGAGGAGCTATCTTGAACCTTCCGACAGGGACACTCC 1080  
 DB 2114 TGTGGGTATTTCTGTAACCTGTTGAGGAGCTATCTTGAACCTTCCGACAGGGACACTCC 2173  
 QY 1081 GCCAGGCCAGGGGTCAAGGGGAGAGTGAACCTGATGATGAGTGAAGTGAAGTGAAGTGA 1140  
 DB 2174 GCCAGGCCAGGGGTCAAGGGGAGAGTGAACCTGATGATGAGTGAAGTGAAGTGAAGTGA 2233  
 QY 1141 GGAGCAGTGTGTTGAGCCTGAGAGCTGTGGGGGAGGAGTGGGGGCTTCTGCTGCT 1200  
 DB 2234 GGAGCAGTGTGTTGAGCCTGAGAGCTGTGGGGGAGGAGTGGGGGCTTCTGCTGCTGCT 2293  
 QY 1201 CATTTGCTTCAATGAAGCTTCAAGAGCCAAACCAAGGCTTTCCCTTCTGAGT 1260  
 DB 2294 CATTTGCTTCAATGAAGCTTCAAGAGCCAAACCAAGGCTTTCCCTTCTGAGT 2353  
 QY 1261 TTGAATATCCAAATCTTTTGTACTTGTGTGTTAAATGTTTAAATTTTGTAAATAAT 1320  
 DB 2354 TTGAATATCCAAATCTTTTGTACTTGTGTGTTAAATGTTTAAATTTTGTAAATAAT 2413  
 QY 1321 AAAATAAATTAATTAATAATGATGTTTCAAGAGCAAACTTCCCT 1368  
 DB 2414 AAAATAAATTAATTAATAATGATGTTTCAAGAGCAAACTTCCCT 2461

RESULT 5  
 ID ABN94640/c  
 ID ABN94640 standard; DNA; 427 BP.  
 XX  
 AC ABN94640;  
 XX  
 DT 13-AUG-2002 (first entry)  
 XX  
 DE Gene #1138 used to diagnose liver cancer.  
 XX  
 KW Gene; liver cancer; ds; hepatocellular carcinoma; hepatotropic;  
 KW metastatic liver tumor; cytostatic; expression profile; disease state;  
 KW disease progression; drug toxicity; drug efficacy; drug metabolism.  
 OS Homo sapiens.  
 XX  
 PN WO200229103-A2.  
 XX  
 PD 11-APR-2002.



```

Db      1109 GGGCTAGCTGAC-----CAGCTGGGCTGCTGCTGACGAGCTGACT 1151
QY      600 TGCACTGGGAGCACTGATGTACCTGGGAACCCCTGGACAGAAAGCTAACCCAGAC 659
Db      1152 -----TGATCCACCTGGGAACCTGGACAGAAAAGCTAGCTCCAGAG 1197
QY      660 AGACAGATGTACAGAGAC-AAACGTGCAATTAATGCAATGTTAAATGTGAGTTTAC 718
Db      1198 CAATGATGTGACAGAGCAAAAGCTGACATATGCAAAATGTTAAATGTGAGTTTGC 1257
QY      719 AGCTTACCTATGAGACTGCTGCTCCTAGTCCAGAAATCAATGGGGGTATGATGCTCTC 778
Db      1258 AGCTT-----CAGTCCCACTGCTGTTGAAACCCCATTTGCTCTC 1296
QY      779 CAACCTGAGGCTGTAAGCA-AGCTCAGGCTAGTCTCCCACTGGGGGCTGAGCCCTC 837
Db      1297 TGAGCTGTAGGCTGTAGAGCAGGCTCAGGCTGCT-----TA 1335
QY      838 CTTGGAGACGGTCTCGTGGCAGCCCATCACTGTGTTCAATGATGAGAAATGACTTAA 897
Db      1336 ACTGGAGACGTCCCGTGGGAG--CCCATTTACTGCAATTG-ATGCTTTGAGAAATGAGCAG 1393
QY      898 AGCCCTGCTGCTGCTGCTGCTGCAATGCAAGCAGAGCGGTGGGGCTGCTGGGAGCAAT 957
Db      1394 AAC-----ACTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1443
QY      958 CCACTGAGAGTCTTCTCAGCTTGTGAGTGTGAGCAGAGAACTTGGCGGGGATGCTCCA 1017
Db      1444 CAGTCTCAGAGTCTCTCTAGTCAAGTCTGCTCCAAATGAGACCGCGGAGTG--CGG 1501
QY      1018 GGAATGAGGATTTCTGTAAGCTGAGGAGGCTATCTGTAAGCTCCGACAGGAGGACTCC 1077
Db      1502 AGATGTGAGTGAACCAAGCACTGAGGAGAGAGGCTCTGGGCTCTCTCTAGAGGTTGCTC 1561
QY      1078 CAGGCAAGCCAGAGGCTCAGAGGAGGAGAGTGAACACTTCAAGTGAAGCCAAAGCTGGGT 1137
Db      1562 TAGGCGAGCCCGAGGCGCGGAGAGAGTGTGCTGATCATATGAGCCAAAGACTAGAGT 1621
QY      1138 CAGGAGAGAGTGTGATTTGAGCCAGCACTGGGAGGAGGAGGAGGAGGAGGAGGAGGAG 1197
Db      1622 GAGGAGAGAGTGTGATTTGAGCCAGCA-CTGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 1680
QY      1198 CCTCATTTGCTTCAATGAAGGCTCAAGAGCAGC-----1232
Db      1681 CCTCATTTGCTTCAATGAAGGCTCAAGAGCAGC-----1232
QY      1233 -----AAAAAGAGGCTTTCCCTCTCTCTGAGTTGAATATC 1269
Db      1741 AGGCAAGGCTCTCCCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 1800
QY      1270 CAGAACTTTTGTACTTCTTGTGTTAATTTGTTAATTTTGTAAATAAATAAATAAATAA 1329
Db      1801 CAGAACTTTTGTACTTCTTGTGTTAATTTGTTAATTTTGTAAATAAATAAATAAATAA 1859
QY      1330 TTAGTTAATTAATTAATTAATTT 1350
Db      1860 TCAATTAATTAATTAATTAATTT 1880

```

## RESULT 7

AA578838 standard; cDNA; 1837 BP.

AA578838;

13-FEB-2002 (first entry)

DNA encoding novel human diagnostic protein #14642.

Human; chromosome mapping; gene mapping; gene therapy; forensic; food supplement; medical imaging; diagnostic; genetic disorder; ss.

```

OS      Homo sapiens.
XX
XX      MO200175067-A2.
XX
XX      11-OCT-2001.
XX
XX      30-MAR-2001; 2001WO-US08631.
XX
XX      31-MAR-2000; 2000US-0540217.
XX
XX      23-AUG-2000; 2000US-0649167.
XX
XX      (HYSE-) HYSEQ INC.
XX
XX      Drmanac RT, Liu C, Tang YT;
XX
XX      WPI; 2001-639362/73.
XX
XX      P-PDB; ABG14651.
XX
XX      New isolated polynucleotide and encoded polypeptides, useful in
XX      diagnostics, forensics, gene mapping, identification of mutations
XX      responsible for genetic disorders or other traits and to assess
XX      biodiversity.
XX
XX      Claim 1; SEQ ID No 14642; 103pp; English.
XX
XX      The invention relates to isolated polynucleotide (I) and
XX      polypeptide (II) sequences. (I) is useful as hybridisation probes,
XX      polymerase chain reaction (PCR) primers, oligomers, and for chromosome
XX      and gene mapping, and in recombinant production of (II). The
XX      polynucleotides are also used in diagnostics as expressed sequence tags
XX      for identifying expressed genes. (I) is useful in gene therapy techniques
XX      to restore normal activity of (II) or to treat disease states involving
XX      (II). (II) is useful for generating antibodies against it, detecting or
XX      quantitating a polypeptide in tissue, as molecular weight markers and as
XX      a food supplement. (II) and its binding partners are useful in medical
XX      imaging of sites expressing (II). (I) and (II) are useful for treating
XX      disorders involving aberrant protein expression or biological activity.
XX      The polypeptide and polynucleotide sequences have applications in
XX      diagnostics, forensics, gene mapping, identification of mutations in
XX      responsible for genetic disorders or other traits to assess biodiversity
XX      and to produce other types of data and products dependent on DNA and
XX      amino acid sequences. AA564197-AA594564 represent novel human
XX      diagnostic coding sequences of the invention.
XX      Note: The sequence data for this patent did not appear in the printed
XX      specification, but was obtained in electronic format directly from WIPO
XX      at ftp.wipo.int/pub/published_pct_sequences.
XX
XX      Sequence 1837 BP; 424 A; 540 C; 471 G; 402 T; 0 other;
XX
XX      Query Match      13.8%; Score 192.8; DB 23; Length 1837;
XX      Best Local Similarity 97.5%; Pred. No. 4,3e-35;
XX      Matches 228; Conservative 0; Mismatches 2; Indels 4; Gaps 4;

```

RESULT 8

AAH86413/c  
ID AAH86413 standard; DNA: 150 BP.

AAH86413;

27-FEB-2002 (first entry)

Human single nucleotide polymorphism containing DNA sequence #1270.

Biallelic marker; polymorphism; human; disease; diagnosis; treatment; phenotypic trait; gene therapy; forensic; paternity; mapping; cancer; transgenic; single nucleotide polymorphism; SNP; ss.

Homo sapiens.

Key Location/Qualifiers  
FT replace(25,T)  
FT /tag= a  
FT /standard\_name= "single nucleotide polymorphism"

MO9953095-A2.

21-OCT-1999.

30-MAR-1999; 99WO-US06893.

09-APR-1998; 98US-0057871.

(WHED) WHITEHEAD INST-BIOMEDICAL RES.

Lander ES, Wang D, Hudson T;

WPI; 1999-620443/53.

Polymorphic human genomic sequences and related allele-specific probes and primers, useful for genetic analysis, e.g. diagnosis and monitoring of disease

Claim 1; Page 167; 330pp; English.

This invention describes novel human nucleic acid segments (I) containing polymorphic sites. The polymorphisms (or disease susceptibility) or e.g. correlating disease polymorphisms (or disease susceptibility) or other phenotypic traits (e.g. baldness, obesity, fertility, strength, response to drugs etc.); diagnosing and monitoring e.g. cancer, inflammation, heart or central nervous system diseases; detecting such susceptibility to microbial infection; treating or preventing such diseases; forensic analysis; gene therapy; paternity testing; mapping genomic loci associated with phenotypic traits (and subsequent cloning of the genes responsible); and the production of transgenic organisms. Antibodies raised against (I) are useful as diagnostic and therapeutic tools and in drug screening. AAH85144 - AAH87644 represent the human DNA sequences containing biallelic polymorphic sites described in the invention.

Sequence 150 BP; 52 A; 19 C; 30 G; 48 T; 1 other;

Query Match 10.7%; Score 149; DB 20; Length 150;  
Best Local Similarity 99.3%; Pred. No. 2.9e-25;  
Matches 149; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

1219 GCCTCAAGACGACCAAAACAGGCTTCCCTCCGAGTTGAATATCCAGATCTT 1278

150 GCCTCAAGACGACCAAAACAGGCTTCCCTCCGAGTTGAATATCCAGATCTT 91

1279 TTGTACTCTCTGTGGTAAATGTTATTTTGTAAAAATTAATAATTAAGTTAAT 1338

90 TTGTACTCTCTGTGGTAAATGTTATTTTGTAAAAATTAATAATTAAGTTAAT 31

1339 AAAATGATGTTTCAAGCAAACTCTCCCT 1368

Db 30 AAAATGATGTTTCAAGCAAACTCTCCCT 1

RESULT 9  
AAH86414/c  
ID AAH86414 standard; DNA: 150 BP.

AAH86414;

27-FEB-2002 (first entry)

Human single nucleotide polymorphism containing DNA sequence #1271.

Biallelic marker; polymorphism; human; disease; diagnosis; treatment; phenotypic trait; gene therapy; forensic; paternity; mapping; cancer; transgenic; single nucleotide polymorphism; SNP; ss.

Homo sapiens.

Key Location/Qualifiers  
FT replace(92,T)  
FT /tag= a  
FT /standard\_name= "single nucleotide polymorphism"

MO9953095-A2.

21-OCT-1999.

30-MAR-1999; 99WO-US06893.

09-APR-1998; 98US-0057871.

(WHED) WHITEHEAD INST BIOMEDICAL RES.

Lander ES, Wang D, Hudson T;

WPI; 1999-620443/53.

Polymorphic human genomic sequences and related allele-specific probes and primers, useful for genetic analysis, e.g. diagnosis and monitoring of disease

Claim 1; Page 167; 330pp; English.

This invention describes novel human nucleic acid segments (I) containing polymorphic sites. The polymorphisms (or disease susceptibility) or e.g. correlating disease polymorphisms (or disease susceptibility) or other phenotypic traits (e.g. baldness, obesity, fertility, strength, response to drugs etc.); diagnosing and monitoring e.g. cancer, inflammation, heart or central nervous system diseases; detecting such susceptibility to microbial infection; treating or preventing such diseases; forensic analysis; gene therapy; paternity testing; mapping genomic loci associated with phenotypic traits (and subsequent cloning of the genes responsible); and the production of transgenic organisms. Antibodies raised against (I) are useful as diagnostic and therapeutic tools and in drug screening. AAH85144 - AAH87644 represent the human DNA sequences containing biallelic polymorphic sites described in the invention.

Sequence 150 BP; 52 A; 19 C; 30 G; 48 T; 1 other;

Query Match 10.7%; Score 149; DB 20; Length 150;  
Best Local Similarity 99.3%; Pred. No. 2.9e-25;  
Matches 149; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

1219 GCCTCAAGACGACCAAAACAGGCTTCCCTCCGAGTTGAATATCCAGATCTT 1278

150 GCCTCAAGACGACCAAAACAGGCTTCCCTCCGAGTTGAATATCCAGATCTT 91

1279 TTGTACTCTCTGTGGTAAATGTTATTTTGTAAAAATTAATAATTAAGTTAAT 1338

90 TTGTACTCTCTGTGGTAAATGTTATTTTGTAAAAATTAATAATTAAGTTAAT 31



QY 1339 AAAATGATGTTTCACAGCAAACTCTTCCT 1368  
DB 30 AAAATGATGTTTCACAGCAAACTCTTCCT 1  
RESULT 10  
ABV05669/c  
ID ABV05669 standard; cDNA; 429 BP.  
AC ABV05669;  
XX  
XX  
DT 13-SEP-2002 (first entry)  
XX  
DE Human prostate expression marker cDNA 5660.  
XX  
XX Human; prostate cancer; cytostatic; carcinogen; pharmacodynamic marker;  
KM pharmacogenomic marker; gene; ss.  
XX  
OS Homo sapiens.  
XX  
PN MO200160860-A2.  
XX  
PD 23-AUG-2001.  
XX  
PF 20-FEB-2001; 2001WO-US05171.  
XX  
PR 17-FEB-2000; 2000US-183319P.  
PR 16-MAR-2000; 2000US-189862P.  
PR 25-MAY-2000; 2000US-207454P.  
PR 09-JUN-2000; 2000US-211314P.  
PR 18-JUL-2000; 2000US-219007P.  
PR 13-DEC-2000; 2000US-255281P.  
XX  
PA (MILL-) MILLENNIUM PREDICTIVE MEDICINE INC.  
XX  
PI Schlegel R, Endege WO, Monahan JE;  
XX  
DR WPI; 2001-662795/76.  
XX  
PT Novel isolated nucleic acid molecule associated with cancerous state of  
PT prostate cells and correlating with presence of prostate cancer, useful  
PT for detecting presence of prostate cancer, stage of prostate cancer  
XX  
XX Claim 1; Page 948; 11750pp; English.  
XX  
XX The invention relates to an isolated nucleic acid molecule (I) comprising  
XX a nucleotide sequence given in Tables 1-9 (ABV00010-ABV62213) of the  
XX specification or its complement. (I) is useful for:  
XX (a) assessing whether a patient is afflicted with prostate cancer;  
XX (b) monitoring the progression of prostate cancer in a patient;  
XX (c) assessing the efficacy of a test compound to inhibit prostate  
XX cancer in a patient;  
XX (d) assessing the efficacy of a therapy for inhibiting prostate cancer  
XX in a patient;  
XX (e) selecting a composition for inhibiting prostate cancer in a patient;  
XX (f) assessing the prostate cell carcinogenic potential of a compound;  
XX (g) determining whether prostate cancer has metastasized in a patient;  
XX (h) assessing the aggressiveness or indolence of prostate cancer in a  
XX patient;  
XX (I) is also useful as a pharmacodynamic or pharmacogenomic marker.  
XX  
SQ Sequence 429 BP; 121 A; 68 C; 61 G; 106 T; 73 other;  
XX  
XX  
Query Match 4.2%; Score 59; DB 23; Length 429;  
Best Local Similarity 48.2%; Pred. No. 0.00043;  
Matches 107; Conservative 0; Mismatches 115; Indels 0; Gaps 0;  
QY 1169 GGGCGCGGGGTGGCGCGGCGCTTCTGCTCATTTGCTTCAATGAAGCCTCAAGC 1228  
DB 259 GNGGNGNCNTTNNAGNTNTTNCCTGGGNGNTCTTGGGNTTTTNNCNCNNAANTN 200  
QY 1229 AGCCAAAACCAAGGCTTCCCTTCGAGTTGAATATCCAGAAATCTTTTACTTCT 1288

DB 199 AGNTTTTTTTTNGNGCCCCCTTGTTNGGTTTTCCTCNCNNAAATTTTNTNGNTTT 140  
QY 1289 TGTGGTAAATGTTTATTTTGTAAATAAATAATAGTTAAATGAATGATG 1348  
DB 139 TTTCCTTTAANTTTTTTTTTTNGNCACAAAAAATTTTTTTTTTTTTTTTTT 80  
QY 1349 TTCACAGCAAACTCTTCCTTAAAAAATTTTTTTTTTTTTTTTTT 1390  
DB 79 AA 38  
RESULT 11  
AAS78836  
ID AAS78836 standard; cDNA; 370 BP.  
AC AAS78836;  
XX  
XX  
DT 13-FEB-2002 (first entry)  
XX  
DE DNA encoding novel human diagnostic protein #14640.  
XX  
XX Human; chromosome mapping; gene mapping; gene therapy; forensic;  
KM food supplement; medical imaging; diagnostic; genetic disorder; ss.  
XX  
OS Homo sapiens.  
XX  
PN MO200175067-A2.  
XX  
PD 11-OCT-2001.  
XX  
PF 30-MAR-2001; 2001WO-US08631.  
XX  
PR 31-MAR-2000; 2000US-0540217.  
PR 23-AUG-2000; 2000US-0649167.  
XX  
XX (HYSE-) HYSEQ INC.  
XX  
PA Drmanac RT, Liu C, Tang YT;  
XX  
PI  
XX  
DR WPI; 2001-639362/73.  
XX  
DR P-PSDB; ABG14649.  
XX  
XX  
PT New isolated polynucleotide and encoded polypeptides, useful in  
PT diagnostics, forensic, gene mapping, identification of mutations  
PT responsible for genetic disorders or other traits and to assess  
PT biodiversity  
XX  
XX Claim 1; SEQ ID No 14640; 103pp; English.  
XX  
XX The invention relates to isolated polynucleotide (I) and  
XX polypeptide (II) sequences. (I) is useful as hybridisation probe,  
XX polynucleotide chain reaction (PCR) primers, oligomers, and for chromosome  
XX and gene mapping, and in recombinant production of (II). The  
XX polynucleotides are also used in diagnostics as expressed sequence tags  
XX for identifying expressed genes. (I) is useful in gene therapy techniques  
XX to restore normal activity of (II) or to treat disease states involving  
XX (II). (II) is useful for generating antibodies against it, detecting or  
XX quantitating a polypeptide in tissue, as molecular weight markers and as  
XX a food supplement. (II) and its binding partners are useful in medical  
XX imaging of sites expressing (II). (I) and (II) are useful for treating  
XX disorders involving aberrant protein expression or biological activity.  
XX The polypeptide and polynucleotide sequences have applications in  
XX diagnostics, forensics, gene mapping, identification of mutations  
XX responsible for genetic disorders or other traits to assess biodiversity  
XX and to produce other types of data and products dependent on DNA and  
XX amino acid sequences. AAS64197-AAS94564 represent novel human  
XX diagnostic coding sequences of the invention.  
XX Note: The sequence data for this patent did not appear in the printed  
XX at ftp.wipo.int/pub/published\_pat\_sequences.  
XX  
SQ Sequence 370 BP; 67 A; 126 C; 101 G; 76 T; 0 other;



	Query Match	3.9%	Score 54.6;	DB 23;	Length 370;	
	Best Local Similarity	93.4%	Pred. No. 0.0043;			
	Matches 57;	Conservative	0;	Mismatches 4;	Indels 0;	Gaps 0;
OY	87	CCATCCCCCAGGAACACTACCGCTGTCGTGGCATCTTACCACCA CGGAGCTGCCTCTTT	146			
Db	309	CACGCCACAGGAAGACTACCGCTGTCGTGGCATCTTACCACCA CGGAGCTGCCTCTTT	368			
OY	147 C 147					
Db	369 C 369					
	RESULT 12					
ID	ABV14838/c					
XX	ABV14838 standard; CDNA; 400 BP.					
XX						
AC	ABV14838;					
XX						
DT	13-SEP-2002 (first entry)					
XX						
DE	Human prostate expression marker CDNA 14829.					
XX						
KX	Human; prostate cancer; cytostatic; carcinogen; pharmacodynamic marker;					
XX	pharmacogenomic marker; gene; ss.					
OS	Homo sapiens.					
PN	WO200160860-A2.					
XX						
PD	23-AUG-2001.					
XX						
PF	20-FEB-2001; 2001WO-US05171.					
XX						
PR	17-FEB-2000; 2000US-183319P.					
XX	16-MAR-2000; 2000US-189862P.					
PR	25-MAY-2000; 2000US-207454P.					
XX	09-JUN-2000; 2000US-211314P.					
PR	18-JUL-2000; 2000US-219007P.					
XX	13-DEC-2000; 2000US-255281P.					
PA	(MILL-) MILLENNIUM PREDICTIVE MEDICINE INC.					
XX						
PI	Schlegel R, Endege WO, Monahan JE;					
XX						
DR	WPI; 2001-662795/76.					
XX						
PT	Novel isolated nucleic acid molecule associated with cancerous state of					
PT	prostate cells and correlating with presence of prostate cancer, useful					
PT	for detecting presence of prostate cancer, stage of prostate cancer -					
PS	Claim 1; Page 2482; 11750pp; English.					
XX						
CC	The invention relates to an isolated nucleic acid molecule (I) comprising					
CC	a nucleotide sequence given in Tables 1-9 (ABV00010-ABV62213) of the					
CC	specification or its complement. (I) is useful for:					
CC	(a) assessing whether a patient is afflicted with prostate cancer;					
CC	(b) monitoring the progression of prostate cancer in a patient;					
CC	(c) assessing the efficacy of a test compound to inhibit prostate					
CC	cancer in a patient;					
CC	(d) assessing the efficacy of a therapy for inhibiting prostate cancer					
CC	in a patient;					
CC	(e) selecting a composition for inhibiting prostate cancer in a patient;					
CC	(f) assessing the prostate cell carcinogenic potential of a compound;					
CC	(g) determining whether prostate cancer has metastasized in a patient;					
CC	(h) assessing the aggressiveness or indolence of prostate cancer in a					
CC	patient;					
CC	(I) is also useful as a pharmacodynamic or pharmacogenomic marker.					
XX						
Sequence	400 BP; 101 A; 77 G; 82 G; 140 T; 0 other;					
Query Match	3.7%; Score 51; DB 23; Length 400;					
Best Local Similarity	56.1%; Pred. No. 0.031;					

Matches 96; Conservative 0; Mismatches 75; Indels 0; Gaps 0;

QY 1220 CCTAAGACGCCAAAACGAGCTTCCCTTCTGAGTTGAATATCCAGAACTCTT 1279  
Db 179 CCCCATAATGATTTTTTTTGGGAGGCCCTGTTGGTTTTTCCCTCCGAAATTTTG 120  
QY 1280 TGTACTCTTGTTGGTAAATTTGTTATTTTTTGTAAATAAATAATTAATTAAGTTAATA 1339  
Db 119 GTAAAGTTTTTCTTTAAATTTTTTTTTTTTGGCCCAAAAAAAAAAAAAAAAAAAAA 60  
QY 1340 AAAAGATGTTTCACAGCAACTCTTCCCTVAAAAAAAAAAAAAAAAAAAA 1390  
Db 59 AA 9

RESULT.13  
ABV38926/C  
ID ABV38926 standard; CDNA; 436 BP.  
XX  
AC ABV38926;  
XX  
DT 16-SEP-2002 (first entry)  
XX  
DE Human prostate expression marker CDNA 38917.  
XX  
KM Human; prostate cancer; cytostatic; carcinogen; pharmacodynamic marker;  
XX pharmacogenomic marker; gene; ss.  
OS Homo sapiens.  
PN WO200160860-A2.  
PD 23-AUG-2001.  
XX  
PF 20-FEB-2001; 2001WO-US05171.  
XX  
PR 17-FEB-2000; 2000US-183319P.  
PR 16-MAR-2000; 2000US-189862P.  
PR 25-MAY-2000; 2000US-207454P.  
PR 09-JUN-2000; 2000US-211314P.  
PR 18-JUL-2000; 2000US-219007P.  
PR 13-DEC-2000; 2000US-255281P.  
XX  
PA (MILL-) MILLENNIUM PREDICTIVE MEDICINE INC.  
XX  
PI Schlegel R, Endege WO, Nonahan JE;  
XX  
DR WPI; 2001-662795/76.  
XX  
XX Novel isolated nucleic acid molecule associated with cancerous state of  
PT prostate cells and correlating with presence of prostate cancer, useful  
PT for detecting presence of prostate cancer, stage of prostate cancer -  
XX  
XX Claim 1; Page 7910-7911, 11750pp; English.  
XX  
XX The invention relates to an isolated nucleic acid molecule (I) comprising  
CC a nucleotide sequence given in Tables 1-9 (ABV00010-ABV62213) of the  
CC specification or its complement. (I) is useful for:  
CC (a) assessing whether a patient is afflicted with prostate cancer;  
CC (b) monitoring the progression of prostate cancer in a patient;  
CC (c) assessing the efficacy of a test compound to inhibit prostate  
CC cancer in a patient;  
CC (d) assessing the efficacy of a therapy for inhibiting prostate cancer  
CC in a patient;  
CC (e) selecting a composition for inhibiting prostate cancer in a patient;  
CC (f) assessing the prostate cell carcinogenic potential of a compound;  
CC (g) determining whether prostate cancer has metastasized in a patient;  
CC (h) assessing the aggressiveness or indolence of prostate cancer in a  
CC patient;  
CC (I) is also useful as a pharmacodynamic or pharmacogenomic marker.  
XX  
XX Sequence 436 BP; 104 A; 89 C; 96 G; 147 T; 0 other;  
XX

Query Match 3.7%; Score 51; DB 23; Length 436;  
Best Local Similarity 56.1%; Pred. No. 0.032;  
Matches 96; Conservative 0; Mismatches 75; Indels 0; Gaps 0;

Db 1220 CCTCAAGCAGCCAAACAGGCTTCCCTCTCGATTTGAATATCCAGAACTTT 1279  
217 CCCCAATAGATTTTCTTTTGGGAGCCCTTGTGTTTCCCTCCCAATTTTG 158  
Qy 1280 TGTACTCTCTGTTGTTAAATGTTTATTTTGAATAAATAAATTAAGTTAATA 1339  
157 GTAGGTTTCTTTCTTTAAATTTTGTGGCCCAAAAAAAAAAAAAAAAAA 98  
Db 1340 AAATGATGTTTCACAGCAACTCTTCCCTAATAAAAAAAAAAAAAA 1390  
97 AA 47

RESULT 14  
ABV44714/c  
ID ABV44714 standard; cDNA; 436 BP.  
AC ABV44714;  
XX  
DT 16-SEP-2002 (first entry)  
XX  
DE Human prostate expression marker cDNA 44705.  
XX  
KW Human; prostate cancer; cytostatic; carcinogen; pharmacodynamic marker;  
KW pharmacogenomic marker; gene; ss.  
XX  
OS Homo sapiens.  
XX  
PN WO200160860-A2.  
XX  
PD 23-AUG-2001.  
XX  
PE 20-FEB-2001; 2001WO-US05171.  
XX  
PF 17-FEB-2000; 2000US-183319P.  
PR 16-MAR-2000; 2000US-189862P.  
PR 25-MAY-2000; 2000US-207454P.  
PR 09-JUN-2000; 2000US-211314P.  
PR 18-JUL-2000; 2000US-219007P.  
PR 13-DEC-2000; 2000US-255281P.  
XX  
PA (MILL-) MILLENNIUM PREDICTIVE MEDICINE INC.  
XX  
PI Schlegel R, Endege WO, Monahan JE;  
XX  
DR WPI; 2001-662795/76.  
XX  
PT Novel isolated nucleic acid molecule associated with cancerous state of  
PT prostate cells and correlating with presence of prostate cancer, useful  
PT for detecting presence of prostate cancer, stage of prostate cancer -  
XX  
PS Claim 1; Page 8866; 11750P; English.  
XX  
CC The invention relates to an isolated nucleic acid molecule (I) comprising  
CC a nucleotide sequence given in Tables 1-9 (ABV00010-ABV62213) of the  
CC specification or its complement. (I) is useful for:  
CC (a) assessing whether a patient is afflicted with prostate cancer;  
CC (b) monitoring the progression of prostate cancer in a patient;  
CC (c) assessing the efficacy of a test compound to inhibit prostate  
CC cancer in a patient;  
CC (d) assessing the efficacy of a therapy for inhibiting prostate cancer  
CC in a patient;  
CC (e) selecting a composition for inhibiting prostate cancer in a patient;  
CC (f) assessing the prostate cell carcinogenic potential of a compound;  
CC (g) determining whether prostate cancer has metastasized in a patient;  
CC (h) assessing the aggressiveness or indolence of prostate cancer in a  
CC patient;  
CC (I) is also useful as a pharmacodynamic or pharmacogenomic marker.  
XX

Sequence 436 BP; 104 A; 89 C; 96 G; 147 T; 0 other;

Query Match 3.7%; Score 51; DB 23; Length 436;  
Best Local Similarity 56.1%; Pred. No. 0.032;  
Matches 96; Conservative 0; Mismatches 75; Indels 0; Gaps 0;

Db 1220 CCTCAAGCAGCCAAACAGGCTTCCCTCTCGATTTGAATATCCAGAACTTT 1279  
217 CCCCAATAGATTTTCTTTTGGGAGCCCTTGTGTTTCCCTCCCAATTTTG 158  
Qy 1280 TGTACTCTCTGTTGTTAAATGTTTATTTTGAATAAATAAATTAAGTTAATA 1339  
157 GTAGGTTTCTTTCTTTAAATTTTGTGGCCCAAAAAAAAAAAAAAAAAA 98  
Db 1340 AAATGATGTTTCACAGCAACTCTTCCCTAATAAAAAAAAAAAAAA 1390  
97 AA 47

RESULT 15  
ABQ67093/c  
ID ABQ67093 standard; DNA; 83391 BP.  
AC ABQ67093;  
XX  
DT 28-AUG-2002 (first entry)  
XX  
DE Human angiogenesis associated polynucleotide SEQ ID NO 123.  
XX  
KW Human; angiogenesis; methylation; eye disease; glaucoma; tumour;  
KW inflammation; rheumatoid arthritis; diabetic retinopathy; antileucers;  
KW macular degeneration; inflammatory bowel disease; Crohn's disease;  
KW antirheumatic; antiarthritic; antidiabetic; antipsoriatic;  
KW arteriosclerotic; ds.  
XX  
OS Homo sapiens.  
XX  
PN WO200246454-A2.  
XX  
PD 13-JUN-2002.  
XX  
PF 06-DEC-2001; 2001WO-BP14320.  
XX  
PR 06-DEC-2000; 2000DE-1061338.  
XX  
PA (EPIC-) EPIGENOMICS AG.  
XX  
PI Schacht O;  
XX  
DR WPI; 2002-500450/53.  
XX  
PT New nucleic acid fragments from chemically treated  
PT angiogenesis-associated genes, useful for determining methylation  
PT status, e.g. in diagnosis or treatment of cancer -  
XX  
PS Claim 1; SEQ ID NO 123; 41pp + Sequence Listing; German.  
XX  
CC The invention relates to a nucleic acid (I) comprising a segment of 18  
CC bases of chemically pretreated DNA of angiogenesis-associated genes (II)  
CC having sequences (ABQ67093-ABQ67178) or their complements. (I), also  
CC related oligomers, are used to evaluate the methylation status and/or  
CC single-nucleotide polymorphisms, in angiogenesis-related genes, for  
CC diagnosis and treatment of eye diseases, proliferative retinopathy,  
CC neovascular glaucoma, solid tumours, inflammation, rheumatoid arthritis,  
CC diabetic retinopathy, macular degeneration caused by neovascularisation,  
CC psoriasis, arteriosclerosis, inflammatory bowel diseases, ulcers and  
CC Crohn's disease.  
CC Note: The sequence data for this patent did not form part of the printed  
CC specification, but was obtained in electronic format directly from WIPO  
CC at [ftp.wipo.int/pub/published\\_pct\\_sequences](http://wipo.int/pub/published_pct_sequences).  
XX  
SQ Sequence 83391 BP; 25113 A; 665 C; 16761 G; 40835 T; 17 other;  
XX

Query Match 3.7%; Score 51; DB 24; Length 83391;  
 Best Local Similarity 54.5%; Pred. No. 0.21;  
 Matches 102; Conservative 0; Mismatches 85; Indels 0; Gaps 0;

QY	1204	TTGCTTCAATGAAGCCCAAGCCCAAGCCCAAGCCCTTCCCTTCCTGAGTTG	1263
DB	31892	TTCTTACATCTAAATCTAATAAACACCCCAACCCCAATTTTAAACCTA	31833
QY	1264	AATATCCAGAACTTTGTAATCTTGTGTTAAATGTTATTTTGTAAAAATAA	1323
DB	31832	AAAAATATTAATTTTATTAATCTTAACCTAATAATAATAATTAATCTTTTACAAAT	31773
QY	1324	ATAAATTAATTAATAAATGATGTTTCAGCAAACTTCCTTAAAAAATAA	1383
DB	31772	ACCTATTTCTTAATAAATAAATAACATATTTTAACTAATAAATAATAA	31713
QY	1384	AAAAAA	1390
DB	31712	AAAAAA	31706

Search completed: June 29, 2003, 04:14:17  
 Top time : 251 secs

GenCore version 5.1.6  
Copyright (c) 1993 - 2003 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: June 29, 2003, 03:52:18 ; Search time 58 Seconds  
(without alignments)  
7386.680 Million cell updates/sec

Title: US-09-989-919-15

Perfect score: 1397  
Sequence: 1 ggtgtctgcacccgtaccgga.....aaaaaaaaaaggcgcgtc 1397

Scoring table: IDENTITY NUC  
Gapop 10.0 Gapext 1.0

Searched: 441362 seqs, 153338381 residues

Total number of hits satisfying chosen parameters: 882724

Minimum DB seq length: 0  
Maximum DB seq length: 200000000

Post-Processing: Minimum Match 0%  
Maximum Match 100%

Listing first 45 summaries

Database : Issued Patents NA:  
1: /cgn2\_6/prodata/1/ina/5A\_COMB.seq:\*  
2: /cgn2\_6/prodata/1/ina/5B\_COMB.seq:\*  
3: /cgn2\_6/prodata/1/ina/5A\_COMB.seq:\*  
4: /cgn2\_6/prodata/1/ina/5B\_COMB.seq:\*  
5: /cgn2\_6/prodata/1/ina/PCTUS\_COMB.seq:\*  
6: /cgn2\_6/prodata/1/ina/backfilseq.seq:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

# SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	43.2	3.1	2223	1 US-08-257-073-4	Sequence 4, Appl1
2	43	3.1	944	3 US-08-306-769-82	Sequence 82, Appl1
3	43	3.1	944	3 US-08-306-616-82	Sequence 82, Appl1
4	43	3.1	944	3 US-08-817-795-82	Sequence 82, Appl1
5	43	3.1	944	3 US-08-639-075A-82	Sequence 82, Appl1
6	43	3.1	944	4 US-09-012-431-82	Sequence 82, Appl1
7	43	3.1	944	4 US-09-012-692-82	Sequence 82, Appl1
8	43	3.1	944	4 US-08-306-613-82	Sequence 82, Appl1
9	43	3.1	944	5 PCT-US95-1442A-82	Sequence 82, Appl1
10	43	3.1	945	4 US-09-032-215-26	Sequence 26, Appl1
11	43	3.1	1582	3 US-08-545-196B-10	Sequence 28, Appl1
12	43	3.1	1582	3 US-08-545-196B-10	Sequence 10, Appl1
13	43	3.1	1582	3 US-08-545-196B-12	Sequence 12, Appl1
14	42.6	3.0	7218	1 US-07-867-106-2	Sequence 2, Appl1
15	42	3.0	5852	1 US-07-867-106-2	Sequence 1, Appl1
16	41.8	3.0	4203	2 US-08-866-753-1	Sequence 1, Appl1
17	41.8	3.0	4203	2 US-08-866-753-1	Sequence 1, Appl1
18	41.4	3.0	721	4 US-09-288-143-38	Sequence 38, Appl1
19	41.4	3.0	6152	4 US-08-973-462-1	Sequence 1, Appl1
20	41.4	3.0	37950	4 US-09-338-907-183	Sequence 183, App
21	41.4	3.0	37950	4 US-09-338-907-183	Sequence 183, App
22	40.8	2.9	198	1 US-08-330-108-16	Sequence 16, Appl1
23	40.8	2.9	198	5 PCT-US92-10087-16	Sequence 16, Appl1
24	40.4	2.9	144	1 US-08-702-344-26	Sequence 26, Appl1
25	40.4	2.9	1474	4 US-08-821-994-64	Sequence 64, Appl1
26	40.4	2.9	3138	1 US-07-867-106-4	Sequence 4, Appl1
27	40	2.9	1683	4 US-09-347-803-11	Sequence 11, Appl1

28	40	2.9	2447	2 US-09-014-969-14	Sequence 14, Appl1
29	39.8	2.8	1129	4 US-09-227-357-40	Sequence 40, Appl1
30	39.6	2.8	1114	4 US-09-152-060-41	Sequence 41, Appl1
31	39.6	2.8	2323	4 US-09-149-476-24	Sequence 24, Appl1
32	39.4	2.8	1558	1 US-08-455-550-7	Sequence 7, Appl1
33	39.4	2.8	1813	4 US-09-071-224-3	Sequence 3, Appl1
34	39.2	2.8	1159	4 US-09-410-464-14	Sequence 14, Appl1
35	39	2.8	1454	4 US-09-372-422A-19	Sequence 19, Appl1
36	39	2.8	2202	4 US-09-465-558-59	Sequence 59, Appl1
37	38.8	2.8	1096	4 US-09-461-697-26	Sequence 26, Appl1
38	38.8	2.8	1147	1 US-08-665-716-1	Sequence 1, Appl1
39	38.6	2.8	1882	4 US-09-370-253-1	Sequence 1, Appl1
40	38.4	2.7	1181	4 US-09-149-476-310	Sequence 310, App
41	38.4	2.7	1212	4 US-09-149-476-186	Sequence 186, App
42	38.4	2.7	2334	1 US-08-062-632-4	Sequence 4, Appl1
43	38.2	2.7	1134	3 US-09-248-335-29	Sequence 29, Appl1
44	38.2	2.7	1492	4 US-08-745-995A-25	Sequence 25, Appl1
45	38.2	2.7	1492	4 US-08-745-995A-27	Sequence 27, Appl1

## ALIGNMENTS

RESULT 1  
US-08-257-073-4  
; Sequence 4, Application US/08257073  
; Patent No. 5766597  
; GENERAL INFORMATION:  
; APPLICANT: Paolelli, Enzo  
; APPLICANT: de Taisne, Charles  
; APPLICANT: Tine, John A.  
; TITLE OF INVENTION: MALARIA RECOMBINANT POXVIRUS VACCINE  
; NUMBER OF SEQUENCES: 143  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Curtis, Morris & Safford, P.C.  
; STREET: 530 Fifth Avenue, 25th Floor  
; CITY: New York  
; STATE: New York  
; COUNTRY: UNITED STATES OF AMERICA  
; ZIP: 10036  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: Patentin Release #1.0, Version #1.30  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/257,073  
; FILING DATE: 09-JUN-1994  
; CLASSIFICATION: 424  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: US 08/075,783  
; FILING DATE: 11-JUN-1993  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: US 07/852,305  
; FILING DATE: 18-MAR-1992  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: US 07/672,183  
; FILING DATE: 20-MAR-1991  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Frommer, William S.  
; REGISTRATION NUMBER: 25,506  
; REFERENCE/DOCKET NUMBER: 454310-2570  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: (212) 840-3333  
; TELEFAX: (212) 840-0712  
; TELEX: 425066 CURTMS  
; INFORMATION FOR SEQ ID NO: 4:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 2223 base pairs  
; TYPE: nucleic acid  
; STRANDEDNESS: single  
; TOPOLOGY: linear  
; US-08-257-073-4

Query Match 3.1%; Score 43.2; DB 1; Length 2223;

Best Local Similarity 63.5%; Pred. No. 0.039; Mismatches 38; Indels 0; Gaps 0;

Matches 66; Conservative 0; Mismatches 38; Indels 0; Gaps 0;

1286 TCTTGTGTTAATGTTTATTTTGTAAATAATAAATAGTTAATAAATGA 1345

1488 TGTATGATGAAATGATTTCTTTGAAAAAGATTTATCGAAAAATATATACACTTA 1547

1346 TGTTCACAGCAACTCTTCTTAAAAAATAAAAA 1389

1548 TGTGTACACCACTCAATCAAAAAAATGAAA 1591

### RESULT 2

US-08-906-769-82

Sequence 82, Application US/08906769

Patent No. 6077687

GENERAL INFORMATION:

APPLICANT: Grieve, Robert B.

APPLICANT: Rushlow, Keith E.

APPLICANT: Wu Hunter, Shirley

APPLICANT: Frank, Glenn R.

APPLICANT: Stiegler, Gary

APPLICANT: Gaines, Patrick J.

APPLICANT: Silver, Gary

TITLE OF INVENTION: FLEA PROTEASE PROTEINS, NUCLEIC ACID

TITLE OF INVENTION: MOLECULES AND USES THEREOF

NUMBER OF SEQUENCES: 190

CORRESPONDENCE ADDRESS:

ADDRESSEE: Sheridan Ross & McIntosh

STREET: 1700 Lincoln Street, Suite 3500

CITY: Denver

STATE: Colorado

COUNTRY: USA

ZIP: 80203

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk

COMPUTER: IBM PC compatible

OPERATING SYSTEM: PC-DOS/MS-DOS

SOFTWARE: Patent in Release #1.0, Version #1.30

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/08/906,769

FILING DATE: 24-APR-1996

CLASSIFICATION:

PRIOR APPLICATION DATA:

APPLICATION NUMBER: US 08/639,075

FILING DATE: 24-APR-1996

ATTORNEY/AGENT INFORMATION:

NAME: Connell, Gary J.

REGISTRATION NUMBER: 32,020

REFERENCE/DOCKET NUMBER: 2618-25-C2

TELECOMMUNICATION INFORMATION:

TELEPHONE: (303) 863-9700

TELEFAX: (303) 863-0223

INFORMATION FOR SEQ. ID NO: 82:

SEQUENCE CHARACTERISTICS:

LENGTH: 944 base pairs

TYPE: nucleic acid

STRANDEDNESS: single

TOPOLOGY: linear

MOLECULE TYPE: cDNA

FEATURE:

NAME/KEY: CDS

LOCATION: 3..768

US-08-906-769-82

Query Match 3.1%; Score 43; DB 3; Length 944;

Best Local Similarity 58.0%; Pred. No. 0.027; Mismatches 55; Indels 0; Gaps 0;

Matches 76; Conservative 0; Mismatches 55; Indels 0; Gaps 0;

1260 TTGAATATCAGAACTTTGTTACTCTTGTGTTAAATGTTTATTTGTAATAA 1319

874 TTATTCACCAAGATGTATTAAGTAAATTAACAATTAATATGTTGTCACAAAAA 933

1380 AAAAAAAAAA 1390

814 TTTAGTATATAAATCCCTTGATGATCAGCAAAATATTTGTTATTTATTTATTC 873

1320 TAAAAAATTAATAGTTAATAATATGTTTACAGCAAACTCTCCCTAAAAA 1379

874 TTATTCACCAAGATGTATTAAGTAAATTAACAATTAATATGTTGTCACAAAAA 933

1380 AAAAAAAAAA 1390

934 AAAAAAAAAA 944

### RESULT 3

US-08-906-616-82

Sequence 82, Application US/08906616

Patent No. 6121035

GENERAL INFORMATION:

APPLICANT: Grieve, Robert B.

APPLICANT: Rushlow, Keith E.

APPLICANT: Wu Hunter, Shirley

APPLICANT: Frank, Glenn R.

APPLICANT: Stiegler, Gary

APPLICANT: Gaines, Patrick J.

APPLICANT: Silver, Gary

TITLE OF INVENTION: FLEA AMINOPEPTIDASE PROTEINS AND USES THEREOF

NUMBER OF SEQUENCES: 190

CORRESPONDENCE ADDRESS:

ADDRESSEE: Sheridan Ross P.C.

STREET: 1700 Lincoln Street, Suite 3500

CITY: Denver

STATE: Colorado

COUNTRY: USA

ZIP: 80203

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk

COMPUTER: IBM PC compatible

OPERATING SYSTEM: PC-DOS/MS-DOS

SOFTWARE: Patent in Release #1.0, Version #1.30

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/08/906,616

FILING DATE: 05-AUG-1997

CLASSIFICATION: 536

ATTORNEY/AGENT INFORMATION:

NAME: Connell, Gary J.

REGISTRATION NUMBER: 32,020

REFERENCE/DOCKET NUMBER: 2618-25-C2-3

TELECOMMUNICATION INFORMATION:

TELEPHONE: (303) 863-9700

TELEFAX: (303) 863-0223

INFORMATION FOR SEQ. ID NO: 82:

SEQUENCE CHARACTERISTICS:

LENGTH: 944 base pairs

TYPE: nucleic acid

STRANDEDNESS: single

TOPOLOGY: linear

MOLECULE TYPE: cDNA

FEATURE:

NAME/KEY: CDS

LOCATION: 3..768

US-08-906-616-82

Query Match 3.1%; Score 43; DB 3; Length 944;

Best Local Similarity 58.0%; Pred. No. 0.027; Mismatches 55; Indels 0; Gaps 0;

Matches 76; Conservative 0; Mismatches 55; Indels 0; Gaps 0;

1260 TTGAATATCAGAACTTTGTTACTCTTGTGTTAAATGTTTATTTGTAATAA 1319

814 TTTAGTATATAAATCCCTTGATGATCAGCAAAATATTTGTTATTTATTTATTC 873

1320 TAAAAAATTAATAGTTAATAATATGTTTACAGCAAACTCTCCCTAAAAA 1379

874 TTATTCACCAAGATGTATTAAGTAAATTAACAATTAATATGTTGTCACAAAAA 933

1380 AAAAAAAAAA 1390

Db 934 AAAAAAAAAA 944

## RESULT 4

US-08-817-795-82

Sequence 82, Application US/08817795

Patent No. 6139840

GENERAL INFORMATION:

APPLICANT: Grieve, Robert B.

APPLICANT: Rushlow, Keith E.

APPLICANT: Hunter, Shirley Wu

APPLICANT: Frank, Glenn W.

APPLICANT: Health, Andrew W.

APPLICANT: Yamaka, Miles Yamanaka

APPLICANT: Arfsten, Ann

APPLICANT: Dale, Beverly

APPLICANT: Stiegler, Gary

TITLE OF INVENTION: USE OF PROTEASE INHIBITORS AND

TITLE OF INVENTION: PROTEASE VACCINES TO PROTECT ANIMALS FROM FLEA

TITLE OF INVENTION: INFESTATION, AND FLEA PROTEASE PROTEINS, NUCLEIC ACID

TITLE OF INVENTION: MOLECULES, AND USES THEREOF

NUMBER OF SEQUENCES: 119

CORRESPONDENCE ADDRESS:

ADDRESSEE: Sheridan Ross &amp; McIntosh,

STREET: 1700 Lincoln Street, Suite 3500

CITY: Denver

STATE: Colorado

COUNTRY: USA

ZIP: 80203

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk

COMPUTER: IBM PC compatible

OPERATING SYSTEM: PC-DOS/MS-DOS

SOFTWARE: Patentin Release #1.0, Version #1.25

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/08/817,795

CLASSIFICATION:

PRIORITY APPLICATION DATA:

APPLICATION NUMBER: PCT/US95/14442

FILING DATE:

ATTORNEY/AGENT INFORMATION:

NAME: Gary J. Connell

REGISTRATION NUMBER: 32,020

REFERENCE/DOCKET NUMBER:

TELEPHONE: (303) 863-9700

TELEFAX: (303) 863-0223

INFORMATION FOR SEQ ID NO: 82:

SEQUENCE CHARACTERISTICS:

LENGTH: 944 base pairs

TYPE: nucleic acid

STRANDEDNESS: single

TOPOLOGY: linear

MOLECULE TYPE: CDNA

FEATURE:

NAME/KEY: CDS

LOCATION: 3..768

US-08-817-795-82

Query Match

Best Local Similarity: 3.1%; Score 43; DB 3; Length 944;

Matches 76; Conservative 0; Mismatches 55; Indels 0; Gaps 0;

Db 1260 TTTGAATATCCAGATCTTTGTTACTCTTTGTTAAATGTTTATTTTGTAAAAA 1319

Db 814 TTTTGTATTAATTAATCTTTGTTGATTCATGCAAAATATTTGTTTATTTATTTATTC 873

Qy 1320 TAAATATAATAGTAAATTAATAGTGTTCACAGCAACTCTTCCCTAAAAA 1379

Db 874 TTTTATCAACGATGATTAATGAATTAACATTAATAAATGTTAGTGTGCCAAAAA 933

Qy 1380 AAAAAAAAAA 1390

Db 934 AAAAAAAAAA 944

## RESULT 5

US-08-639-075A-82

Sequence 82, Application US/08639075A

Patent No. 6150125

GENERAL INFORMATION:

APPLICANT: Grieve, Robert B.

APPLICANT: Rushlow, Keith E.

APPLICANT: Wu Hunter, Shirley

APPLICANT: Frank, Glenn R.

APPLICANT: Stiegler, Gary

APPLICANT: Gaines, Patrick J.

APPLICANT: Silver, Gary

TITLE OF INVENTION: FLEA PROTEASE PROTEINS, NUCLEIC ACID

TITLE OF INVENTION: MOLECULES AND USES THEREOF

NUMBER OF SEQUENCES: 190

CORRESPONDENCE ADDRESS:

ADDRESSEE: Sheridan Ross &amp; McIntosh,

STREET: 1700 Lincoln Street, Suite 3500

CITY: Denver

STATE: Colorado

COUNTRY: USA

ZIP: 80203

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk

COMPUTER: IBM PC compatible

OPERATING SYSTEM: PC-DOS/MS-DOS

SOFTWARE: Patentin Release #1.0, Version #1.30

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/08/639,075A

FILING DATE: 24-APR-1996

CLASSIFICATION: 424

ATTORNEY/AGENT INFORMATION:

NAME: Connell, Gary J.

REGISTRATION NUMBER: 32,020

REFERENCE/DOCKET NUMBER: 2618-25-C2

TELEPHONE: (303) 863-9700

TELEFAX: (303) 863-0223

INFORMATION FOR SEQ ID NO: 82:

SEQUENCE CHARACTERISTICS:

LENGTH: 944 base pairs

TYPE: nucleic acid

STRANDEDNESS: single

TOPOLOGY: linear

MOLECULE TYPE: CDNA

FEATURE:

NAME/KEY: CDS

LOCATION: 3..768

US-08-639-075A-82

Query Match

Best Local Similarity: 3.1%; Score 43; DB 3; Length 944;

Matches 76; Conservative 0; Mismatches 55; Indels 0; Gaps 0;

Db 1260 TTTGAATATCCAGATCTTTGTTACTCTTTGTTAAATGTTTATTTTGTAAAAA 1319

Db 814 TTTTGTATTAATTAATCTTTGTTGATTCATGCAAAATATTTGTTTATTTATTTATTC 873

Qy 1320 TAAATATAATAGTAAATTAATAGTGTTCACAGCAACTCTTCCCTAAAAA 1379

Db 874 TTTTATCAACGATGATTAATGAATTAACATTAATAAATGTTAGTGTGCCAAAAA 933

Qy 1380 AAAAAAAAAA 1390

Db 934 AAAAAAAAAA 944

## RESULT 6

US-09-012-431-82  
Sequence 82, Application US/09012431  
Patent No. 6180383  
GENERAL INFORMATION:  
APPLICANT: Griewe, Robert B.  
Rushlow, Keith E.  
Wu Hunter, Shirley  
Frank, Glenn R.  
Stiegler, Gary  
Gaines, Patrick J.  
Silver, Gary  
TITLE OF INVENTION: FLEA PROTEASE PROTEINS, NUCLEIC ACID  
MOLECULES AND USES THEREOF  
NUMBER OF SEQUENCES: 190  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Sheridan Rose & McIntosh  
STREET: 1700 Lincoln Street, Suite 3500  
CITY: Denver  
STATE: Colorado  
COUNTRY: USA  
ZIP: 80203  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patentin Release #1.0, Version #1.30  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/09/012,431  
FILING DATE: 23-Jan-1998  
CLASSIFICATION: <Unknown>  
PRIORITY APPLICATION DATA:  
APPLICATION NUMBER: US 08/639,075  
FILING DATE: <Unknown>  
ATTORNEY/AGENT INFORMATION:  
NAME: Connell, Gary J.  
REGISTRATION NUMBER: 32,020  
REFERENCE/DOCKET NUMBER: 2618-25-C2  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (303) 863-9700  
TELEFAX: (303) 863-0223  
INFORMATION FOR SEQ ID NO: 82:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 944 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
MOLECULE TYPE: cDNA  
FEATURE:  
NAME/KEY: CDS  
LOCATION: 3..768  
SEQUENCE DESCRIPTION: SEQ ID NO: 82:  
US-09-012-431-82  
Query Match 3.1%; Score 43; DB 4; Length 944;  
Best Local Similarity 58.0%; Pred. No. 0.027; Indels 0; Gaps 0;  
Matches 76; Conservative 0; Mismatches 55;  
Db 1260 TTGAATATCCAGAACTTTTGTACTCTTGTGGTTAAATGTTTATTTTGTAAAAA 1319  
814 TTTAGATATAAATCTTTGATTCATGCAAAATTTTGTATTATTTATTTAC 873  
QY 1320 TAAATATAATAGTATAAATGATGTTTCACAGCAACTCTCCCTAAAAA 1379  
Db 874 TTTATTCAGAAAGTATTAAGTAATTAACAATAAATGTTAGTGTCCAAAAA 933  
QY 1380 AAAAAAAAAA 1390  
Db 934 AAAAAAAAAA 944

RESULT 7  
US-09-012-692-82  
Sequence 82, Application US/09012692

Patent No. 6214579  
GENERAL INFORMATION:  
APPLICANT: Griewe, Robert B.  
Rushlow, Keith E.  
Wu Hunter, Shirley  
Frank, Glenn R.  
Stiegler, Gary  
Gaines, Patrick J.  
Silver, Gary  
TITLE OF INVENTION: FLEA PROTEASE PROTEINS, NUCLEIC ACID  
MOLECULES AND USES THEREOF  
NUMBER OF SEQUENCES: 190  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Sheridan Rose & McIntosh  
STREET: 1700 Lincoln Street, Suite 3500  
CITY: Denver  
STATE: Colorado  
COUNTRY: USA  
ZIP: 80203  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patentin Release #1.0, Version #1.30  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/09/012,692  
FILING DATE:  
CLASSIFICATION:  
PRIORITY APPLICATION DATA:  
APPLICATION NUMBER: US 08/639,075  
FILING DATE: 24-APR-1996  
ATTORNEY/AGENT INFORMATION:  
NAME: Connell, Gary J.  
REGISTRATION NUMBER: 32,020  
REFERENCE/DOCKET NUMBER: 2618-25-C2  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (303) 863-9700  
TELEFAX: (303) 863-0223  
INFORMATION FOR SEQ ID NO: 82:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 944 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
MOLECULE TYPE: cDNA  
FEATURE:  
NAME/KEY: CDS  
LOCATION: 3..768  
US-09-012-692-82  
Query Match 3.1%; Score 43; DB 4; Length 944;  
Best Local Similarity 58.0%; Pred. No. 0.027; Indels 0; Gaps 0;  
Matches 76; Conservative 0; Mismatches 55;  
Db 1260 TTGAATATCCAGAACTTTTGTACTCTTGTGGTTAAATGTTTATTTTGTAAAAA 1319  
814 TTTAGATATAAATCTTTGATTCATGCAAAATTTTGTATTATTTATTTAC 873  
QY 1320 TAAATATAATAGTATAAATGATGTTTCACAGCAACTCTCCCTAAAAA 1379  
Db 874 TTTATTCAGAAAGTATTAAGTAATTAACAATAAATGTTAGTGTCCAAAAA 933  
QY 1380 AAAAAAAAAA 1390  
Db 934 AAAAAAAAAA 944

RESULT 8  
US-08-906-613-82  
Sequence 82, Application US/08906613  
Patent No. 6232096  
GENERAL INFORMATION:  
APPLICANT: Griewe, Robert B.

APPLICANT: Rushlow, Keith E.  
APPLICANT: Wu Hunter, Shirley  
APPLICANT: Frank, Glenn R.  
APPLICANT: Stiegler, Gary  
APPLICANT: Gaines, Patrick J.  
APPLICANT: Silver, Gary  
TITLE OF INVENTION: FLEA PROTEASE PROTEINS, NUCLEIC ACID  
TITLE OF INVENTION: MOLECULES AND USES THEREOF  
NUMBER OF SEQUENCES: 190  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Sheridan Ross & McIntosh  
STREET: 1700 Lincoln Street, Suite 3500  
CITY: Denver  
STATE: Colorado  
COUNTRY: USA  
ZIP: 80203  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patentin Release #1.0, Version #1.30  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/906,613  
FILING DATE:  
CLASSIFICATION:  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 08/639,075  
FILING DATE: 24-Apr-1996  
ATTORNEY/AGENT INFORMATION:  
NAME: Connell, Gary J.  
REGISTRATION NUMBER: 32,020  
REFERENCE/DOCKET NUMBER: 2618-25-C2  
TELEPHONE: (303) 863-9700  
TELEFAX: (303) 863-0223  
INFORMATION FOR SEQ ID NO: 82:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 944 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
MOLECULE TYPE: cDNA  
FEATURE:  
NAME/KEY: CDS  
LOCATION: 3..768  
US-08-906-613-82

Query Match 3.1%; Score 43; DB 4; Length 944;  
Best Local Similarity 58.0%; Pred. No. 0.027;  
Matches 76; Conservative 0; Mismatches 55; Indels 0; Gaps 0;

QY 1260 TTGGATATCCAGAAATTTTGTGACTCTTGTTGGTTAAATGTTTATTTTGTAAATAA 1319  
DB 814 TTTTATGATTAATAATCTTTGATTCATGCAAAATATTTGTTTATTTATTTTATTTAC 873  
QY 1320 TAAATTAATAATGTTAATAATGATGTTTCACAGCAACTCTTCCCTAAATAAATAA 1379  
DB 874 TTTATTCAAACGATGATTAAGTGAATTAACATATAAATGTTAGTGTGCCAAAAA 933  
QY 1380 AAAAAAAAAA 1390  
DB 934 AAAAAAAAAA 944

RESULT 9  
PCT-US95-14442A-82  
Sequence 82, Application PC/TUS9514442A  
GENERAL INFORMATION:  
APPLICANT: Grieve, Robert B.  
APPLICANT: Rushlow, Keith E.  
APPLICANT: Hunter, Shirley Wu  
APPLICANT: Frank, Glenn R.  
APPLICANT: Heath, Andrew W.

APPLICANT: Yamaka, Miles Yamanaka  
APPLICANT: Arfsten, Ann  
APPLICANT: Dale, Beverly  
APPLICANT: Stiegler, Gary  
TITLE OF INVENTION: USE OF PROTEASE INHIBITORS AND  
TITLE OF INVENTION: PROTEASE VACCINES TO PROTECT ANIMALS FROM FLEA  
TITLE OF INVENTION: INFESTATION, AND FLEA PROTEASE PROTEINS, NUCLEIC ACID  
NUMBER OF SEQUENCES: 119  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Sheridan Ross & McIntosh  
STREET: 1700 Lincoln Street, Suite 3500  
CITY: Denver  
STATE: Colorado  
COUNTRY: USA  
ZIP: 80203  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patentin Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: PCT/US95/14442A  
FILING DATE:  
CLASSIFICATION:  
ATTORNEY/AGENT INFORMATION:  
NAME: Gary J. Connell  
REGISTRATION NUMBER: 32,020  
REFERENCE/DOCKET NUMBER:  
TELEPHONE: (303) 863-9700  
TELEFAX: (303) 863-0223  
INFORMATION FOR SEQ ID NO: 82:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 944 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
MOLECULE TYPE: cDNA  
FEATURE:  
NAME/KEY: CDS  
LOCATION: 3..768  
PCT-US95-14442A-82

Query Match 3.1%; Score 43; DB 5; Length 944;  
Best Local Similarity 58.0%; Pred. No. 0.027;  
Matches 76; Conservative 0; Mismatches 55; Indels 0; Gaps 0;

QY 1260 TTGGATATCCAGAAATTTTGTGACTCTTGTTGGTTAAATGTTTATTTTGTAAATAA 1319  
DB 814 TTTTATGATTAATAATCTTTGATTCATGCAAAATATTTGTTTATTTATTTTATTTAC 873  
QY 1320 TAAATTAATAATGTTAATAATGATGTTTCACAGCAACTCTTCCCTAAATAAATAA 1379  
DB 874 TTTATTCAAACGATGATTAAGTGAATTAACATATAAATGTTAGTGTGCCAAAAA 933  
QY 1380 AAAAAAAAAA 1390  
DB 934 AAAAAAAAAA 944

RESULT 10  
US-09-032-215-26  
Sequence 26, Application US/09032215  
GENERAL INFORMATION:  
APPLICANT: Stiegler, Gary L.  
APPLICANT: Gaines, Patrick J.  
TITLE OF INVENTION: FLEA PROTEASE PROTEINS, NUCLEIC  
TITLE OF INVENTION: ACID MOLECULES, AND USES THEREOF  
NUMBER OF SEQUENCES: 50  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Sheridan Ross P.C.



STREET: 1700 Lincoln Street, Suite 3500  
CITY: Denver  
STATE: Colorado  
COUNTRY: U.S.A.  
ZIP: 80203  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: ASCII DOS TEXT  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/09/032.215  
FILING DATE: 27-FEB-1998  
CLASSIFICATION: 536  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER:  
FILING DATE:  
ATTORNEY/AGENT INFORMATION:  
NAME: Connell, Gary J.  
REGISTRATION NUMBER: 32,020  
REFERENCE/DOCKET NUMBER: 2618-25-C6  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (303) 863-9700  
TELEFAX: (303) 863-0223  
INFORMATION FOR SEQ ID NO: 26:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 945 nucleotides  
TYPE: nucleic acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
MOLECULE TYPE: CDNA  
FEATURE:  
NAME/KEY: CDS  
LOCATION: 1..768  
US-09-032-215-26

Query Match  
Best Local Similarity 3.1%; Score 43; DB 4; Length 945;  
Matches 76; Conservative 0; Mismatches 55; Indels 0; Gaps 0;

Db 1260 TTGAATATCCAGAACTTTTGTACTCTGTGTGTTAAATGTTATTTTGTAAAAA 1319  
815 TTTTAGTAATAATAATCTTTGTATTCATGCAAAATTTTGTATTATTATTATTTAC 874

QY 1320 TAAATATAATTTAGTTAATAAATGATGTTTACAGCAAACTCTCCCTAAAAAAA 1379  
875 TTTATTCAACGAATGTAATAGTAATTAACAATAAATGTTAGTGTGCCAAAAAA 934

Db 1380 AAAAAAAAAA 1390  
935 AAAAAAAAAA 945

RESULT 11  
US-09-032-215-28/C  
Sequence 28, Application US/09032215  
Patent No. 6204010  
GENERAL INFORMATION:  
APPLICANT: Stiegler, Gary L.  
APPLICANT: Gaines, Patrick J.  
TITLE OF INVENTION: FLEA PROTEASE PROTEINS, NUCLEIC  
TITLE OF INVENTION: ACID MOLECULES, AND USES THEREOF  
NUMBER OF SEQUENCES: 50  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Sheridan, Ross P.C.  
STREET: 1700 Lincoln Street, Suite 3500  
CITY: Denver  
STATE: Colorado  
COUNTRY: U.S.A.  
ZIP: 80203  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible

OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: ASCII DOS TEXT  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/09/032.215  
FILING DATE: 27-FEB-1998  
CLASSIFICATION: 536  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER:  
FILING DATE:  
ATTORNEY/AGENT INFORMATION:  
NAME: Connell, Gary J.  
REGISTRATION NUMBER: 32,020  
REFERENCE/DOCKET NUMBER: 2618-25-C6  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (303) 863-9700  
TELEFAX: (303) 863-0223  
INFORMATION FOR SEQ ID NO: 28:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 945 nucleotides  
TYPE: nucleic acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
MOLECULE TYPE: CDNA  
US-09-032-215-28

Query Match  
Best Local Similarity 3.1%; Score 43; DB 4; Length 945;  
Matches 76; Conservative 0; Mismatches 55; Indels 0; Gaps 0;

Db 1260 TTGAATATCCAGAACTTTTGTACTCTGTGTGTTAAATGTTATTTTGTAAAAA 1319  
131 TTTAGTAATAATAATCTTTGTATTCATGCAAAATTTTGTATTATTATTATTTAC 72

QY 1320 TAAATATAATTTAGTTAATAAATGATGTTTACAGCAAACTCTCCCTAAAAAAA 1379  
71 TTTATTCAACGAATGTAATAGTAATTAACAATAAATGTTAGTGTGCCAAAAAA 12

Db 1380 AAAAAAAAAA 1390  
11 AAAAAAAAAA 1

RESULT 12  
US-08-545-196B-10  
Sequence 10, Application US/08545196B  
Patent No. 6080577  
GENERAL INFORMATION:  
APPLICANT: MELK, JUDITH  
APPLICANT: MUNNICH, ARNOLD  
TITLE OF INVENTION: SURVIVAL MOTOR NEURON (SMN) GENE: A GENE  
TITLE OF INVENTION: FOR SPINAL MUSCULAR ATROPHY  
NUMBER OF SEQUENCES: 65  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: BIRCH, STEWART, KOLASCH AND BIRCH, LLP  
STREET: PO BOX 747  
CITY: FALLS CHURCH  
STATE: VA  
COUNTRY: USA  
ZIP: 22040-0747  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patentin Release #1.0, Version #1.30  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/545.196B  
FILING DATE: 19-OCT-1995  
CLASSIFICATION: 435  
ATTORNEY/AGENT INFORMATION:  
NAME: FARACT, C. J.  
REGISTRATION NUMBER: 32,350  
REFERENCE/DOCKET NUMBER: 2121-110P  
TELECOMMUNICATION INFORMATION:





GenCore version 5.1.6  
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: June 29, 2003, 04:10:13 ; Search time 166 Seconds  
(without alignments)  
12492.616 Million cell updates/sec

Title: US-09-989-919-15

Perfect score: 1397

Sequence: 1 ggtgctgcacgtcaccgga.....aaaaaaaaaagcgcgc 1397

Scoring table: IDENTITY NUC

Gapop 10.0, Gapext 1.0

Searched: 1055720 seqs, 742224136 residues

Total number of hits satisfying chosen parameters: 2111440

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

Published Applications\_NA.\*  
1: /cgn2\_6/prodata/1/pubpna/US07\_PUBCOMB.seq.\*  
2: /cgn2\_6/prodata/1/pubpna/PCT\_NEW\_PUB.seq.\*  
3: /cgn2\_6/prodata/1/pubpna/US06\_NEW\_PUB.seq.\*  
4: /cgn2\_6/prodata/1/pubpna/US06\_PUBCOMB.seq.\*  
5: /cgn2\_6/prodata/1/pubpna/US07\_NEW\_PUB.seq.\*  
6: /cgn2\_6/prodata/1/pubpna/PCTUS\_PUBCOMB.seq.\*  
7: /cgn2\_6/prodata/1/pubpna/US08\_NEW\_PUB.seq.\*  
8: /cgn2\_6/prodata/1/pubpna/US08\_PUBCOMB.seq.\*  
9: /cgn2\_6/prodata/1/pubpna/US09\_NEW\_PUB.seq.\*  
10: /cgn2\_6/prodata/1/pubpna/US09\_PUBCOMB.seq.\*  
11: /cgn2\_6/prodata/1/pubpna/US10\_NEW\_PUB.seq.\*  
12: /cgn2\_6/prodata/1/pubpna/US10\_PUBCOMB.seq.\*  
13: /cgn2\_6/prodata/1/pubpna/US60\_NEW\_PUB.seq.\*  
14: /cgn2\_6/prodata/1/pubpna/US60\_PUBCOMB.seq.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

# SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	1397	100.0	1397	9 US-09-989-919-15	Sequence 15, Appl
2	464.2	33.2	470	9 US-09-989-919-14	Sequence 1138, Ap
3	420.8	30.1	427	10 US-09-980-107-1138	Sequence 1138, Ap
4	97	6.9	493	9 US-09-918-995-32213	Sequence 32213, A
5	55.2	4.0	428	10 US-09-960-352-10180	Sequence 10180, A
6	51	3.7	435	10 US-09-960-352-15014	Sequence 15014, A
7	50.6	3.6	435	10 US-09-960-352-15014	Sequence 15014, A
8	49.4	3.5	435	10 US-09-960-352-15014	Sequence 15014, A
9	49	3.5	361	9 US-10-198-846-8619	Sequence 8619, Ap
10	48.8	3.5	375	10 US-09-960-352-15014	Sequence 15014, A
11	48.8	3.5	424	10 US-09-960-352-11218	Sequence 11218, A
12	48.6	3.5	746	10 US-09-910-943-714	Sequence 714, App
13	48.4	3.5	298	10 US-09-960-352-1004	Sequence 1004, A
14	48.2	3.5	436	10 US-09-834-975-533	Sequence 533, App
15	48	3.4	424	9 US-10-198-846-2929	Sequence 2929, Ap
16	47.8	3.4	277	10 US-09-960-352-12673	Sequence 12673, A
17	47.8	3.4	312	10 US-09-960-352-8414	Sequence 8414, A
18	47.2	3.4	403	10 US-09-960-352-3319	Sequence 3319, Ap
19	47	3.4	638	9 US-10-198-846-8560	Sequence 8560, Ap

C 20	46.8	3.4	600	9 US-10-198-846-8434	Sequence 8434, Ap
C 21	46.8	3.4	1767	12 US-10-001-843-1	Sequence 1, Appl
C 22	46.6	3.3	416	10 US-09-960-352-4584	Sequence 4584, Ap
C 23	46.2	3.3	418	10 US-09-960-352-4845	Sequence 4845, Ap
C 24	46	3.3	242	10 US-09-960-352-3217	Sequence 3217, Ap
C 25	46	3.3	368	10 US-09-960-352-9843	Sequence 9843, Ap
C 26	46	3.3	448	9 US-10-198-846-2540	Sequence 2540, Ap
C 27	46	3.3	469	9 US-09-918-995-13017	Sequence 13017, A
C 28	46	3.3	2773	9 US-09-992-598-178	Sequence 178, App
C 29	46	3.3	2773	9 US-09-989-2928-178	Sequence 178, App
C 30	46	3.3	2773	9 US-10-063-547-33	Sequence 33, Appl
C 31	46	3.3	2773	9 US-09-989-735-178	Sequence 178, App
C 32	46	3.3	2773	9 US-09-990-444-178	Sequence 178, App
C 33	46	3.3	2773	9 US-09-989-730-178	Sequence 178, App
C 34	46	3.3	2773	9 US-09-990-436-178	Sequence 178, App
C 35	46	3.3	2773	9 US-09-991-181-178	Sequence 178, App
C 36	46	3.3	2773	9 US-09-993-687-178	Sequence 178, App
C 37	46	3.3	2773	9 US-09-989-734-178	Sequence 178, App
C 38	46	3.3	2773	9 US-09-997-653-178	Sequence 178, App
C 39	46	3.3	2773	9 US-10-174-590-149	Sequence 149, App
C 40	46	3.3	2773	9 US-10-176-758-149	Sequence 149, App
C 41	46	3.3	2773	9 US-10-063-616-33	Sequence 33, Appl
C 42	46	3.3	2773	9 US-10-175-737-149	Sequence 149, App
C 43	46	3.3	2773	9 US-09-993-667-178	Sequence 178, App
C 44	46	3.3	2773	9 US-10-063-502-33	Sequence 33, Appl
C 45	46	3.3	2773	9 US-10-173-706-149	Sequence 149, App

## ALIGNMENTS

RESULT 1	US-09-989-919-15
Sequence 15, Application US/09989919	
Patent No. US20020164344A1	
GENERAL INFORMATION:	
APPLICANT: Macina, Roberto	
APPLICANT: Recipon, Hervé	
APPLICANT: Pluta, Jason	
APPLICANT: Ghosh, Malavika	
APPLICANT: Sun, Yongming	
APPLICANT: Liu, Chenghua	
TITLE OF INVENTION: Compositions and Methods Relating to Colon Specific Genes and Pr	
FILE REFERENCE: DEX-0289	
CURRENT APPLICATION NUMBER: US/09/989,919	
CURRENT FILING DATE: 2001-11-21	
PRIOR APPLICATION NUMBER: 60/252,505	
PRIOR FILING DATE: 2000-11-22	
NUMBER OF SEQ ID NOS: 124	
SOFTWARE: PatentIn version 3.1	
SEQ ID NO 15	
LENGTH: 1397	
TYPE: DNA	
ORGANISM: Homo sapien	
US-09-989-919-15	
Query Match	100.0%; Score 1397; DB 9; Length 1397;
Best Local Similarity	100.0%; Pred. No. 0;
Matches 1397; Conservative	0; Mismatches 0; Indels 0; Gaps 0;
QY	1 GGTGTCACCTGTATCCAGACAGACATCCCAAGAGACTACCGTGTGGCCATC 60
DB	1 GGTGTCACCTGTATCCAGACAGACATCCCAAGAGACTACCGTGTGGCCATC 60
QY	61 CGAGTACAGTGTATCCAGACAGACATCCCAAGAGACTACCGTGTGGCCATC 120
DB	61 CGAGTACAGTGTATCCAGACAGACATCCCAAGAGACTACCGTGTGGCCATC 120
QY	121 CTAACACACGAGAGAGTGTCTTCACTGTTCACCTGCTGAGAGCTGTGATCTCG 180
DB	121 CTAACACACGAGAGAGTGTCTTCACTGTTCACCTGCTGAGAGCTGTGATCTCG 180
QY	181 TGAAGACCATGCTGCTGCTTGTGTGACCAACAGACACCTGAGCAGGTGA 240

181 TGAGAGCATGCCAGTGTGGGCTTTGTGTGTCACCAACCAACCACTGGAGAGTGA 240  
241 GCCAGTGGAGAAACCTCTTCCAAAGGAGATGGACAGCTCTCTGGAGTTGATAGTAG 300  
241 GCCAGTGGAGAAACCTCTTCCAAAGGAGATGGACAGCTCTCTGGAGTTGATAGTAG 300  
301 TGATCCCCCATCGGAAGTCAAGAGGGGTGTGAGGTATGAGAGAGAGATACGTGTCT 360  
301 TGATCCCCCATCGGAAGTCAAGAGGGGTGTGAGGTATGAGAGAGAGATACGTGTCT 360  
361 TCAAGGAGTCAAAATTGAGAGAAATGTGTCTTGCCTCCAGAAAGAAACATCCAGCCCTG 420  
361 TCAAGGAGTCAAAATTGAGAGAAATGTGTCTTGCCTCCAGAAAGAAACATCCAGCCCTG 420  
421 TTACCTCTCACTTGTGCCCCCAGGTGGAGAGTGTCTTTTTCAGAGCTGATGAGACC 480  
421 TTACCTCTCACTTGTGCCCCCAGGTGGAGAGTGTCTTTTTCAGAGCTGATGAGACC 480  
481 AAGTGTCCCTGATCCCAAGAGCAATATGTGAAGGCTCTGGCTGATCTATCTGAG 540  
481 AAGTGTCCCTGATCCCAAGAGCAATATGTGAAGGCTCTGGCTGATCTATCTGAG 540  
541 GGGCTGGCTGACAGCTGATCTCTGAGAGCTGGGCTTGGCTGTGAGAGAGTGA 600  
541 GGGCTGGCTGACAGCTGATCTCTGAGAGCTGGGCTTGGCTGTGAGAGAGTGA 600  
601 GCACTGGACAGCTGATCTGACCTGGAGAACCCCTGGAGACAAAGCTAACCTCCAGACA 660  
601 GCACTGGACAGCTGATCTGACCTGGAGAACCCCTGGAGACAAAGCTAACCTCCAGACA 660  
661 GACAGATGTACCAAGAGCAAAAGCTGCAATATGCAAAATGTAAATGTAGTTACAG 720  
661 GACAGATGTACCAAGAGCAAAAGCTGCAATATGCAAAATGTAAATGTAGTTACAG 720  
721 CTTAGATGAGAGCTGTGGCTCCTAGTCCAGAAATCATGGGGGTATGACTGCTCTCCA 780  
721 CTTAGATGAGAGCTGTGGCTCCTAGTCCAGAAATCATGGGGGTATGACTGCTCTCCA 780  
781 ACCCTGTGGCTGTAGCAAAAGCTAGTCTCCCACTGGGGGTGTGGCTCTCTCT 840  
781 ACCCTGTGGCTGTAGCAAAAGCTAGTCTCCCACTGGGGGTGTGGCTCTCTCT 840  
841 GGAAGCTTCCGTGGAGAGCCCATCATCTGTGTTCAATAGTGAAGATGTAGTTAAGC 900  
841 GGAAGCTTCCGTGGAGAGCCCATCATCTGTGTTCAATAGTGAAGATGTAGTTAAGC 900  
901 CCTGT 960  
901 CCTGT 960  
961 TGT 1020  
961 TGT 1020  
1021 TGT 1080  
1021 TGT 1080  
1081 GGCAGCCAGAGGGTCAAGAGAGAGTGTGCAACCTGAGCATGAGCCAAAGCTGGGGGTGAG 1140  
1081 GGCAGCCAGAGGGTCAAGAGAGAGTGTGCAACCTGAGCATGAGCCAAAGCTGGGGGTGAG 1140  
1081 GGCAGCCAGAGGGTCAAGAGAGAGTGTGCAACCTGAGCATGAGCCAAAGCTGGGGGTGAG 1140  
1081 GGCAGCCAGAGGGTCAAGAGAGAGTGTGCAACCTGAGCATGAGCCAAAGCTGGGGGTGAG 1140  
1141 GGAAGCAGT 1200  
1141 GGAAGCAGT 1200  
1141 GGAAGCAGT 1200  
1201 CATTTGCTTTCAATGAAGCTCAAGAGCCAAACCAAGCTTTTCCCTCTCTGAGT 1260  
1201 CATTTGCTTTCAATGAAGCTCAAGAGCCAAACCAAGCTTTTCCCTCTCTGAGT 1260  
1201 CATTTGCTTTCAATGAAGCTCAAGAGCCAAACCAAGCTTTTCCCTCTCTGAGT 1260  
1261 TTGAATATCAGAAATCTTTGTACTTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 1320  
1261 TTGAATATCAGAAATCTTTGTACTTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 1320

1261 TTGAATATCAGAAATCTTTGTACTTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 1320  
1321 AAAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 1380  
1321 AAAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 1380  
1381 AAAAAAAAAAGCGGTC 1397  
1381 AAAAAAAAAAGCGGTC 1397

US-09-989-919-14  
; Sequence 14, Application US/09989919  
; Patent No. US2002016434A1  
; GENERAL INFORMATION:  
; APPLICANT: Macina, Roberto  
; APPLICANT: Recipon, Hervé  
; APPLICANT: Pluta, Jason  
; APPLICANT: Ghosh, Malavika  
; APPLICANT: Sun, Yongming  
; APPLICANT: Liu, Chenghua  
; TITLE OF INVENTION: Compositions and Methods Relating to Colon Specific Genes and Peptides  
; FILE REFERENCE: DEX-0289  
; CURRENT APPLICATION NUMBER: US/09/989,919  
; PRIOR FILING DATE: 2001-11-21  
; PRIOR APPLICATION NUMBER: 60/252,505  
; NUMBER OF SEQ ID NOS: 124  
; SOFTWARE: PatentIn version 3.1  
; SEQ ID NO 14  
; LENGTH: 470  
; TYPE: DNA  
; ORGANISM: Homo sapien  
US-09-989-919-14  
Query Match 33.2%; Score 464.2; DB: 9; Length 470;  
Best Local Similarity 99.4%; Pred. No. 36-114;  
Matches 466; Conservative 0; Mismatches 3; Indels 0; Gaps 0;  
567 CAGCAGCTGGGCTTGGCTGTGAGAGGAGTGTGCTGCTGAGCAGCAGCTGATGCTCACTG 626  
2 CAGGCTGGGCTTGGCTGTGAGAGGAGTGTGCTGCTGAGCAGCAGCTGATGCTCACTG 61  
627 GGAACCCCTGACAGCAAAAGCTAATCCAGACAGACAGATGTGACAGAGCAAAAGTGC 686  
62 GGAACCCCTGACAGCAAAAGCTAATCCAGACAGACAGATGTGACAGAGCAAAAGTGC 121  
687 AATTAATGCCAAATGTTAAATGAGTTTACAGAGCTGATGAGAGCTGCTGCTCTA 746  
122 AATTAATGCCAAATGTTAAATGAGTTTACAGAGCTGATGAGAGCTGCTGCTCTA 181  
747 GTCCAGAAATCAATGGGGGTATGACTGCTCTCCAAACCTGTGGCTGTAAAGAGTCA 806  
182 GTCCAGAAATCAATGGGGGTATGACTGCTCTCCAAACCTGTGGCTGTAAAGAGTCA 241  
807 GCTAGTCTCCCACTGGGGGTGTGCTGCTCTCTCTGAGAGGTTCCGTGGCAGCCCATC 866  
242 GCTAGTCTCCCACTGGGGGTGTGCTGCTCTCTCTGAGAGGTTCCGTGGCAGCCCATC 301  
867 ACTGTGTTCAATGATGTGAGAAATGTAGTAAAGCCCTGCTGCTGCTGCTGCAATGCCA 926  
302 ACTGTGTTCAATGATGTGAGAAATGTAGTAAAGCCCTGCTGCTGCTGCTGCAATGCCA 361  
927 CAGCAGCCGTTGGGGCTGCTGTTGGAGACATCCATGTTGAGTGTCTCAAGCTTAA 986  
362 CAGCAGCCGTTGGGGCTGCTGTTGGAGACATCCATGTTGAGTGTCTCAAGCTTAA 421  
987 CTGAGCAGAGAACTTGGCGGGGATGCTCCAGAGATGTGGTATTTGT 1035  
422 CTGAGCAGAGAACTTGGCGGGGATGCTCCAGAGATGTGGTATTTGT 470

RESULT 3  
US-09-880-107-1138/c  
Sequence 1138, Application US/09880107  
Patent No. US20020142981A1  
GENERAL INFORMATION:  
APPLICANT: Horne, Darci T.  
APPLICANT: Vockley, Joseph G.  
APPLICANT: Scherf, Uwe  
APPLICANT: Gene Logic, Inc.  
TITLE OF INVENTION: Gene Expression Profiles in Liver Cancer  
FILE REFERENCE: 44921-5028-WO  
CURRENT APPLICATION NUMBER: US/09/880,107  
CURRENT FILING DATE: 2001-06-14  
PRIOR APPLICATION NUMBER: US 60/211,379  
PRIOR FILING DATE: 2000-06-14  
PRIOR APPLICATION NUMBER: US 60/237,054  
PRIOR FILING DATE: 2000-10-02  
NUMBER OF SEQ ID NOS: 3950  
SOFTWARE: Patentin Ver. 2.1  
SEQ ID NO 1138  
LENGTH: 427  
TYPE: DNA  
ORGANISM: Homo sapiens  
FEATURE:  
OTHER INFORMATION: Genbank Accession No. US20020142981A1 AA451877  
US-09-880-107-1138

Query Match  
Best Local Similarity 30.1%; Score 420.8; DB 10; Length 427;  
Matches 422; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Db  
936 GTGGGGGCTCGCTGGGAGACATCATCGAGAGTCTTCTCAGTTAGGTGACAGG 995  
427 GTGGGGGCTCGCTGGGAGACATCATCGAGAGTCTTCTCAGTTAGGTGACAGG 368  
QY 996 AGACTTGGGGGGGATGCTCCAGAGATGAGGATGCTGTAACCTGGGGAGGCTATCTTG 1055  
Db 367 AGACTTGGGGGGGATGCTCCAGAGATGAGGATGCTGTAACCTGGGGAGGCTATCTTG 308  
QY 1056 ACTTCCCGACAGGGGACATCTCCAGAGCCAGCCAGGGGTCAGGGGACAGAGTGCACACT 1115  
Db 307 ACTTCCCGACAGGGGACATCTCCAGAGCCAGCCAGGGGTCAGGGGACAGAGTGCACACT 248  
QY 1116 CAGCATGAGCAAGACTGGGGGTGAGGAGAGAGGTTGTTGAGCCAGAGCCTGGGGCGG 1175  
Db 247 CAGCATGAGCAAGACTGGGGGTGAGGAGAGAGGTTGTTGAGCCAGAGCCTGGGGCGG 188  
QY 1176 GGGTGGGGCCGGGGGCTTTCTGCTCATTTGCTTCAATGAAGCCTCAAGAGCCAGAA 1235  
Db 187 GGGTGGGGCCGGGGGCTTTCTGCTCATTTGCTTCAATGAAGCCTCAAGAGCCAGAA 128  
QY 1236 ACCAGGCTTCCCTCCCTCCAGGTTGAAATATCCAGAACTTTTGTACTCTTGTGTGT 1295  
Db 127 ACCAGGCTTCCCTCCCTCCAGGTTGAAATATCCAGAACTTTTGTACTCTTGTGTGT 68  
QY 1296 TAAATGTTATTTTGTGAAAAAATAAATAGTAAATGAATGAATGTTTACAG 1355  
Db 67 TAAATGTTATTTTGTGAAAAAATAAATAGTAAATGAATGAATGTTTACAG 8  
QY 1356 CAAA 1359  
Db 7 CAAA 4

RESULT 4  
US-09-918-995-32213  
Sequence 32213, Application US/09918995  
Patent No. US20030073623A1  
GENERAL INFORMATION:  
APPLICANT: Hyseq, Inc.  
TITLE OF INVENTION: NOVEL NUCLEIC ACID SEQUENCES OBTAINED  
FROM VARIOUS CDNA LIBRARIES  
FILE REFERENCE: 20411-756

CURRENT APPLICATION NUMBER: US/09/918,995  
CURRENT FILING DATE: 2001-07-30  
PRIOR APPLICATION NUMBER: US/09/235,076  
PRIOR FILING DATE: 1999-01-20  
NUMBER OF SEQ ID NOS: 38054  
SOFTWARE: FastSeq for Windows Version 3.0  
SEQ ID NO 32213  
LENGTH: 493  
TYPE: DNA  
ORGANISM: Homo sapiens  
FEATURE:  
NAME/KEY: misc.feature  
LOCATION: (1)...(493)  
OTHER INFORMATION: n = A,T,C or G  
US-09-918-995-32213

Query Match  
Best Local Similarity 6.9%; Score 97; DB 9; Length 493;  
Matches 97; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Db  
QY 1 GGTGCTGACCTGTATCCGGAGGGGCGCATCTGACAGAACTCCAGGCAAGCAGCTATC 60  
Db 397 GGTGCTGACCTGTATCCGGAGGGGCGCATCTGACAGAACTCCAGGCAAGCAGCTATC 456  
QY 61 CGAGTACAGTGTATCCAGACAGACATCCCGCAG 97  
Db 457 CGAGTACAGTGTATCCAGACAGACATCCCGCAG 493

RESULT 5  
US-09-960-352-10180/c  
Sequence 10180, Application US/09960352  
Patent No. US20020137139A1  
GENERAL INFORMATION:  
APPLICANT: Warren, Wesley C.  
APPLICANT: Tao, Nengbing  
APPLICANT: Byatt, John C.  
APPLICANT: Mathialagan, Nagappan  
TITLE OF INVENTION: NUCLEIC ACID AND OTHER MOLECULES ASSOCIATED WITH LACTATION AND  
FILE REFERENCE: 16511.006/37-21(10298)C  
CURRENT APPLICATION NUMBER: US/09/960,352  
CURRENT FILING DATE: 2001-09-24  
NUMBER OF SEQ ID NOS: 15112  
SEQ ID NO 10180  
LENGTH: 428  
TYPE: DNA  
ORGANISM: Bos taurus  
OTHER INFORMATION: Clone ID: 44-LIB188-026-Q1-E1-C12  
US-09-960-352-10180

Query Match  
Best Local Similarity 4.0%; Score 55.2; DB 10; Length 428;  
Matches 117; Conservative 0; Mismatches 43; Indels 6; Gaps 3;

Db  
QY 903 CTGCTGCTGCTGCTGACATGACAGAGGGGTGGGGGCTGGCGTGGGGAGCAATCCATC 962  
Db 427 CTGCTGCTGCTGCTGACATGACAGAGGGGTGGGGGCTGGCGTGGGGAGCAATCCATC 370  
QY 963 GTGAGG--TGTCTCTACCTTATGCTGAGAGAGAGTGTGGGGGATGCTCCAGGA 1020  
Db 369 ATGAGAGAGTCTCTTACCAAGCTCCGAGCGAGAGATTGAA--GGATGCTCCAGGA 312  
QY 1021 TGTGGGTATCTGTACTGCGGAGGCTATCTTGACTCCCGACA 1066  
Db 311 TGTGGGTATCTGTACTGCGGAGGCTATCTTGACTCCCGACA 266

RESULT 6  
US-09-960-352-15014  
Sequence 15014, Application US/09960352  
Patent No. US20020137139A1  
GENERAL INFORMATION:

	Query Match	3.6%	Score 50.6	DB 10	Length 425
	Best Local Similarity	53.7%	Pred. No. 0.0021		
	Matches 101	Conservative	0	Mismatches 87	Indels 0
					Gaps 0
QY	1203	TTTGCTTCATGAAAGCCTCAAAACAGCCAAACACAGCTTTCCTCTCGAGTT	1262		
Db	223	TTTTTTTTTTTAAAAAAAAGGGGGGGGAAAAAAACCCTTTTTTTTTTTTTT	164		
QY	1263	GAATATCCAGAACTTTTGTACTTCTGTGGTTAATTGTTATTTTGTAAAAATA	1322		

```

RESULT 9
US-10-198-846-8619/c
Sequence 8619, Application US/10198846
Publication No. US20030099974A1
GENERAL INFORMATION:
APPLICANT: Lillie, James
APPLICANT: Xu, Yongyao
APPLICANT: Wang, Youzhen
APPLICANT: Seelmann, Kathleen
TITLE OF INVENTION: NOVEL GENES, COMPOSITIONS, KITS, AND METHODS
TITLE OF INVENTION: FOR IDENTIFICATION, ASSESSMENT, PREVENTION, AND
TITLE OF INVENTION: THERAPY OF BREAST CANCER
FILE REFERENCE: MRI-049
CURRENT APPLICATION NUMBER: US/10/0198,846
CURRENT FILING DATE: 2002-07-18
PRIOR APPLICATION NUMBER: 60/306,220
PRIOR FILING DATE: 2001-07-18
NUMBER OF SEQ ID NOS: 14084
SOFTWARE: FASTSEQ for Windows Version 4.0
SEQ ID NO 8619
LENGTH: 361
TYPE: DNA
ORGANISM: Homo sapiens

```



FEATURE:  
NAME/KEY: misc feature  
LOCATION: 15, 21, 25, 26, 29, 30, 33, 58, 59, 242, 243, 245, 247, 249,  
LOCATION: 266, 268, 279, 281, 282, 283, 285, 287, 289, 290, 296,  
LOCATION: 297, 311, 312, 313, 324, 325, 326, 329, 331, 343, 344, 351,  
LOCATION: 352, 353, 361  
OTHER INFORMATION: n = A,T,C or G  
US-10-198-846-8619

Query Match 3.5%; Score 49; DB 9; Length 361;  
Best Local Similarity 49.7%; Pred. No. 0.0052;  
Matches 85; Conservative 0; Mismatches 86; Indels 0; Gaps 0;

QY 1220 CCTCAAGCAGCACCAGGCTTCCCTCTCGAGTTGAATATCCAGATCTTT 1279  
DB 335 CCCCGNGGNN 276  
QY 1280 TGTACTCTCTGTTGTTAAATCTTTATTTTGTAAATAATAATTAATTAATTA 1339  
DB 275 TTTTNTNTNN 216  
QY 1340 AATGATGTTTCACGCAACTCTCCCTMAAAAAAAAAAAAAAAAAAAAA 1390  
DB 215 AA 165

## RESULT 10

US-09-960-352-15014/C  
Sequence 15014, Application US/09960352  
Patent No. US20020137139A1  
GENERAL INFORMATION:  
APPLICANT: Warren, Wesley C.  
APPLICANT: Tao, Nengbing  
APPLICANT: Byatt, John C.  
APPLICANT: Mathialagan, Nagappan  
TITLE OF INVENTION: NUCLEIC ACID AND OTHER MOLECULES ASSOCIATED WITH LACTATION AND  
FILE REFERENCE: 16511.006/37-21(10298)C  
CURRENT APPLICATION NUMBER: US/09/960,352  
CURRENT FILING DATE: 2001-09-24  
NUMBER OF SEQ ID NOS: 15112  
SEQ ID NO 15014  
LENGTH: 375  
TYPE: DNA  
ORGANISM: Bos taurus  
OTHER INFORMATION: Clone ID: 64-LIB3058-048-Q1-K1-H8  
US-09-960-352-15014

Query Match 3.5%; Score 48.8; DB 10; Length 375;  
Best Local Similarity 58.1%; Pred. No. 0.006;  
Matches 86; Conservative 0; Mismatches 62; Indels 0; Gaps 0;

QY 1243 TTTCCTCTCTCGAGTTGAATATCCAGATCTTTGACTCTGTTGTTAAATG 1302  
DB 246 TTTTTCATCTCTTTTAAAAAATAATATATATTTTAAAAAATAATATATTT 187  
QY 1303 TTTATTTTGTAAATAATAATAATAATAATAATAATAATAATAATAATA 1362  
DB 186 TTTAAATTAATAATAATAATAATAATAATAATAATAATAATAATAATA 127  
QY 1363 TTCCCTAATAAAAAAAAAAAAAAAAAAAAA 1390  
DB 126 AAAAAAAAAAAAAAAAAAAAAAAAAAAAAA 99

## RESULT 11

US-09-960-352-11218/C  
Sequence 11218, Application US/09960352  
Patent No. US20020137139A1  
GENERAL INFORMATION:  
APPLICANT: Warren, Wesley C.  
APPLICANT: Tao, Nengbing  
APPLICANT: Byatt, John C.

APPLICANT: Mathialagan, Nagappan  
TITLE OF INVENTION: NUCLEIC ACID AND OTHER MOLECULES ASSOCIATED WITH LACTATION AND  
FILE REFERENCE: 16511.006/37-21(10298)C  
CURRENT APPLICATION NUMBER: US/09/960,352  
CURRENT FILING DATE: 2001-09-24  
NUMBER OF SEQ ID NOS: 15112  
SEQ ID NO 11218  
LENGTH: 424  
TYPE: DNA  
ORGANISM: Bos taurus  
OTHER INFORMATION: Clone ID: 48-LIB3058-026-Q1-K1-D12  
US-09-960-352-11218

Query Match 3.5%; Score 48.8; DB 10; Length 424;  
Best Local Similarity 59.3%; Pred. No. 0.0065;  
Matches 83; Conservative 0; Mismatches 57; Indels 0; Gaps 0;

QY 1251 TTCTCGAGTTGAATATCCAGATCTTTGACTCTGTTGTTAAATGTTAATTT 1310  
DB 166 TTTTATTTTATTTTATTTTATTTTATTTTATTTTATTTTATTTTATTTT 107  
QY 1311 TGTAAATAATAATAATAATAATAATAATAATAATAATAATAATAATA 1370  
DB 106 TTAATTTAATAATAATAATAATAATAATAATAATAATAATAATAATA 47  
QY 1371 AAAAAAAAAAAAAAAAAAAAAAAAAAAAA 1390  
DB 46 AAAAAAAAAAAAAAAAAAAAAAAAAAAAA 27

## RESULT 12

US-09-910-943-714  
Sequence 714, Application US/09910943  
Patent No. US20020081610A1  
GENERAL INFORMATION:  
APPLICANT: Hemmati-Briyanlou, Ali  
APPLICANT: Altman, Curtis  
TITLE OF INVENTION: Assays and Materials for Embryonic Gene Expression  
FILE REFERENCE: 7529/16148US1  
CURRENT APPLICATION NUMBER: US/09/910,943  
CURRENT FILING DATE: 2001-07-23  
NUMBER OF SEQ ID NOS: 742  
SOFTWARE: PatentIn version 3.1  
SEQ ID NO 714  
LENGTH: 746  
TYPE: DNA  
ORGANISM: Xenopus laevis  
FEATURE:  
NAME/KEY: misc feature  
LOCATION: (1) (746)  
OTHER INFORMATION: n may be a or g or c or t/u  
US-09-910-943-714

Query Match 3.5%; Score 48.6; DB 10; Length 746;  
Best Local Similarity 56.2%; Pred. No. 0.011; 70; Indels 0; Gaps 0;  
Matches 90; Conservative 0; Mismatches 57; Indels 0; Gaps 0;

QY 1238 CAGGCTTCCCTCTCTCGAGTTGAATATCCAGATCTTTGACTCTGTTGTTA 1297  
DB 125 CAGGATATTTCCCTCCCTCATTTGTTTCCAAAGAAATTTTGTCTTATCTTGAT 184  
QY 1298 AATGTTTATTTTGTAAATAATAATAATAATAATAATAATAATAATAATA 1357  
DB 185 AACTTTTATTAATATCAATCAAAAAAAAAAAAAAAAAAAAAAAAAAAAAA 244  
QY 1358 AACTCTCCCTAATAAAAAAAAAAAAAAAAAAAAAAGGCGTC 1397  
DB 245 AAAAAAAAAAAAAAAAAAAAAAAAAAAAAAGGCGGC 284

## RESULT 13

US-09-960-352-1004/C



	Matches	77;	Conservative	0;	Mismatches	49;	Indels	0;	Gaps	0;
Qy	1271	AGAACTTTTGACTCTTGTTGGTTAAATGTTATTTTGTATAAAATATAAATAAT	1338							
Db	143	AGGCACTTTTATCTCTTTTGGATTAAATATTTGGCTGTCAAAAAAAAAAAAA	84							

Search completed: June 29, 2003, 05:25:37  
Job time: 168 secs

Search completed: June 29, 2003, 05:25:37  
Job time : 168 secs

GenCore version 5.1.6  
Copyright (c) 1993 - 2003 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: June 29, 2003, 02:17:52 ; Search time 1458 Seconds

(without alignments)  
15517.898 Million cell updates/sec

Title: US-09-989-919-15

Perfect score: 1397

Sequence: 1 ggtgctgcacccctaccgga.....aaaaaaaaaaagcgctc 1397

Scoring table: IDENTITY NUC

Gapop 10.0, Gapext 1.0

Searched: 16154066 seqs, 8097743376 residues

Total number of hits satisfying chosen parameters: 32308132

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : EST.\*

1: em\_estda.\*  
2: em\_esthum.\*  
3: em\_estin.\*  
4: em\_estnu.\*  
5: em\_estov.\*  
6: em\_estpl.\*  
7: em\_estro.\*  
8: em\_hic.\*  
9: gb\_est1.\*  
10: gb\_est2.\*  
11: gb\_hic.\*  
12: gb\_est3.\*  
13: gb\_est4.\*  
14: gb\_est5.\*  
15: em\_estfun.\*  
16: em\_estcom.\*  
17: gb\_gss.\*  
18: em\_gss\_hum.\*  
19: em\_gss\_inv.\*  
20: em\_gss\_pln.\*  
21: em\_gss\_vrt.\*  
22: em\_gss\_fun.\*  
23: em\_gss\_mam.\*  
24: em\_gss\_mus.\*  
25: em\_gss\_other.\*  
26: em\_gss\_pro.\*  
27: em\_gss\_rtd.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

# SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	703.2	50.3	1039	14	BM927766 AGENCOURT
2	643.6	46.1	1038	14	BM927767 AGENCOURT
3	626.4	44.8	773	14	BM671384 UI-E-CKI-
4	605.8	43.4	917	14	BM654344 AGENCOURT
5	602.8	43.1	645	13	BM655930 UI-E-DXI-
6	593	42.4	610	13	BM668538 UI-E-CKI-

C 7	588.4	42.1	678	10	AM409897	AM409897 fh03a06.y
8	574.6	41.1	776	12	BM827019	BM827019 602749053
9	562.6	40.3	770	10	BE250262	BE250262 600943273
C 10	562.4	40.3	565	10	AM516922	AM516922 xp86g09.x
11	554	39.7	935	14	BM647652	BM647652 AGENCOURT
12	545.4	39.0	548	14	BM692817	BM692817 UI-E-CKI-
C 13	541.4	38.8	543	13	BM142660	BM142660 1926c02.x
C 14	533.4	38.2	823	14	BM888182	BM888182 AGENCOURT
C 15	528	37.8	529	9	AI669816	AI669816 tui31f08.x
C 16	520	37.2	520	14	BM013425	BM013425 UI-E-CKI-
C 17	519.4	37.2	522	10	AM592430	AM592430 hf42c03.x
C 18	500.4	35.8	504	10	BM677292	BM677292 7083b01.x
C 19	499.8	35.8	504	9	AI808803	AI808803 w197d10.x
C 20	497	35.6	518	9	AI313395	AI313395 qp80a02.x
C 21	495.2	35.4	888	14	BM652861	BM652861 AGENCOURT
C 22	494.8	35.4	498	10	AM512747	AM512747 xt77a02.x
C 23	492.4	35.2	521	14	BM680899	BM680899 UI-E-EJO-
C 24	491.4	35.2	493	10	BE205791	BE205791 bb49b01.x
C 25	490.8	35.1	494	9	AA552700	AA552700 nk15f10.s
C 26	488.8	35.0	492	9	AI453409	AI453409 t37h12.x
C 27	487.2	34.9	543	14	BM726293	BM726293 UI-E-EJO-
C 28	481.4	34.5	999	14	BM068521	BM068521 AGENCOURT
C 29	479.4	34.3	481	10	BM027861	BM027861 w862b06.x
C 30	478.4	34.2	480	9	AI357244	AI357244 qx63b03.x
C 31	477.8	34.2	504	10	AM411057	AM411057 fh10b03.y
C 32	475.8	34.1	490	9	AI581877	AI581877 as02e05.x
C 33	475.6	34.0	665	12	BF528878	BF528878 602643341
C 34	475.4	34.0	489	10	AM001523	AM001523 wu33h08.x
C 35	473	33.9	485	14	BM012938	BM012938 UI-E-CKI-
C 36	465	33.3	465	14	BM693225	BM693225 UI-E-CKI-
C 37	461.8	33.1	507	9	AA037424	AA037424 2K36c04.s
C 38	458	32.8	800	12	BM826158	BM826158 602750515
C 39	457.4	32.7	459	10	AA263720	AA263720 xq47e02.x
C 40	457.2	32.7	563	9	AA135720	AA135720 z111b08.s
C 41	452.4	32.4	454	10	AM058133	AM058133 wv83f07.x
C 42	450.4	32.2	452	9	AI694266	AI694266 wu11g08.x
C 43	449.2	32.2	845	13	BM041039	BM041039 603644776
C 44	446.4	32.0	458	10	AM151856	AM151856 xf72g11.x
C 45	437.4	31.3	497	9	AA129876	AA129876 z113c09.s

## ALIGNMENTS

RESULT 1  
LOCUS BM927766 1039 bp mRNA linear EST 12-MAR-2002  
DEFINITION AGENCOURT 6729770 NIH\_MGC\_100 Homo sapiens CDNA clone IMAGE:5797598  
5', mRNA sequence.  
ACCESSION BM927766  
VERSION BM927766.1 GI:19378145  
KEYWORDS EST.  
SOURCE human.  
ORGANISM Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.  
REFERENCE 1 (bases 1 to 1039)  
NIH-MGC <http://mgc.nci.nih.gov/>  
TITLE National Institutes of Health, Mammalian Gene Collection (MGC)  
JOURNAL Unpublished (1999)  
COMMENT Contact: Robert Strausberg, Ph.D.  
Email: cgabbs-remail.nih.gov  
Tissue Procurement: CGAP (Stanford)  
CDNA Library Preparation: Rubin Laboratory  
CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LNL)  
DNA Sequencing by: Agencourt Bioscience Corporation  
Clone distribution: MGC clone distribution information can be  
found through the I.M.A.G.E. Consortium/LNL at:  
<http://image.jnl.gov>  
Plate: LNCM2024 row: d column: 15  
High quality sequence stop: 655.  
Location/Qualifiers  
1..1039

FEATURES  
source

```

/organism="Homo sapiens"
/db xref="taxon:9606"
/clone_id="IMAGE:5797598"
/lab host="NIH_MGC_100"
/tissue_type="hepatocellular carcinoma, cell line"
/notes="Organ: liver; Vector: pORF7; Site 1: XhoI; Site 2: EcoRI; cDNA made by oligo-dT priming. Directionally cloned into EcoRI/XhoI sites using the following 5' adaptor: GGCACGAG(G). Size-selected >500bp for average insert size 1.8kb. Library constructed by Ling Hong in the laboratory of Gerald M. Rubin (University of California, Berkeley) using ZAP-cDNA synthesis kit (Stratagene) and Superscript II RT (Life Technologies). Note: this is a NIH_MGC Library."
BASE COUNT      223 a      276 c      300 g      237 t      3 others
ORIGIN
Query Match      50.3%; Score 703.2; DB 14; Length 1039;
Best Local Similarity 86.6%; Pred. No. 4.7e-109;
Matches 836; Conservative 0; Mismatches 111; Indels 18; Gaps 5;

420 GTTACCTCTGACCTGCCCCCAGGTGCGACGCTGCTTTTTCAGACTGATGAGC 479
72 GTACACCAACGACCACTGAGACGTCGGAGCTGCTTTTTCAGACTGATGAGC 131
480 CAAGTGTCTCTGATCCCAACAGACCAATATGTAAAGCCTCTGCTGATCTATGTA 539
132 CAAGTGTCTCTGATCCCAACAGACCAATATGTAAAGCCTCTGCTGATCTATGTA 191
540 GGGCTGGGCTGACGAGTGACTTCTCAGAGCTGGGCTTGGCTGTGAGGGAGTACT 599
192 GGGCTGGGCTGACGAGTGACTTCTCAGAGCTGGGCTTGGCTGTGAGGGAGTACT 251
600 TGCATGCGACGACATGTCATGTCACCTGGGAACCCCTGACAGAAAGCTAACCTCCAGAC 659
252 TGCATGCGACGACATGTCATGTCACCTGGGAACCCCTGACAGAAAGCTAACCTCCAGAC 311
660 AGACATGTGACGACGACCAAACTGCAATATGCAATATGCAATATGCAATATGCA 719
312 AGACATGTGACGACGACCAAACTGCAATATGCAATATGCAATATGCAATATGCA 371
720 GCGTAGTATGAGGACTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 779
372 GCGTAGTATGAGGACTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 431
780 AACCTGTGGGCTGTAAAGCAAGCTCAGGCTAGTCTCCCACTGGGGCTGTGCCCTGCC 839
432 AACCTGTGGGCTGTAAAGCAAGCTCAGGCTAGTCTCCCACTGGGGCTGTGCCCTGCC 491
840 TGGGACGCTTCCGCTGGGACGCCCATCATCTGTCTTCAATGTGCAATATGCAATATG 899
492 TGGGACGCTTCCGCTGGGACGCCCATCATCTGTCTTCAATGTGCAATATGCAATATG 551
900 CCCCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 959
552 CCCCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 611
960 ATCGTGAAGTGTCTTCAAGCTTGAAGTGTGACAGAGACTTGGCGGGGAGTGTCCAGG 1019
612 ATCGTGAAGTGTCTTCAAGCTTGAAGTGTGACAGAGACTTGGCGGGGAGTGTCCAGG 671
1020 ATGTGGGTATTTCTGTAACCTGCGGAGGCTATCTTGAACCTTCCCAAGGGGACACTCCCA 1079
672 ATGTGGGTATTTCTGTAACCTGCGGAGGCTATCTTGAACCTTCCCAAGGGGACACTCCCA 731
1080 GGCAGCAGCC--AGGGGTGAGGGGAGAGGTGACACCTGACAGCATGAGCAGAGTGGAGT 1137
732 NGCCAGCCCAAGGGGTGAGGGGAGAGGTGACACCTGACAGCATGAGCAGAGTGGAGT 791
1138 -----CAGGAGAGAGGTGTGTGAGCCAGAGACTGGGGCGGGGGTGGGGCGGG 1188
792 TCAGGAACCAAGGTGTGTGAGCCAGAGACTGGGGCGGGGGTGGGGCGGGCGGGG 851

```

```

1189 GCGTTCTG--CCTCATTTCTTTCAA---TGAAGCTCAAGCAGCCAAACAGGCTT 1244
852 CCTTTCGCGCTCATTTCTTTCATTAATAAAACCTTCAAGCAGCCAAACAGGCTT 911
1245 TCCCGCTTCTC--GAGTTGAATATTCAGAACTTTTGTACTCTTGTGTTAAATT 1301
912 TCCCGCTTCTC--GAGTTGAATATTCAGAACTTTTGTACTCTTGTGTTAAATT 971
1302 GTTTATTTTGAATAAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAA 1361
972 AAAATATGTTATTTTGTGTAACAAATAAATAAATAAATAAATAAATAAATAAATAA 1031
1362 CTTCC 1366
1032 TCCCC 1036

RESULT 2
BM927767      1038 bp      mRNA      linear      EST 12-MAR-2002
LOCUS      BM927767
DEFINITION      AGENCOURT 6729786 NIH_MGC_100 Homo sapiens cDNA clone IMAGE:5797599
5', mRNA sequence.
ACCESSION      BM927767
VERSION      BM927767.1 GI:19378146
KEYWORDS      EST.
SOURCE      human.
ORGANISM      Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homindae; Homo.
REFERENCE      NIH-MGC http://img.nci.nih.gov/.
1 (bases 1 to 1038)
NATIONAL INSTITUTES OF HEALTH, Mammalian Gene Collection (MGC)
UNPUBLISHED (1999)
CONTACT: Robert Strausberg, Ph.D.
Email: cgabbs-remail.nih.gov
Tissue Procurement: CGAP (Stanford)
CDNA Library Preparation: Rubin Laboratory
DNA Sequencing by: Agencourt Bioscience Corporation
Clone distribution: MGC clone distribution information can be
found through the I.M.A.G.E. Consortium/LLNL at:
http://image.llnl.gov
Plate: LLCM2024 row: d column: 16
High quality sequence stop: 476.
Location/Qualifiers
1. 1038
/organism="Homo sapiens"
/db xref="taxon:9606"
/clone_id="IMAGE:5797599"
/lab host="NIH_MGC_100"
/tissue_type="hepatocellular carcinoma, cell line"
/notes="Organ: liver; Vector: pORF7; Site 1: XhoI; Site 2: EcoRI; cDNA made by oligo-dT priming. Directionally cloned into EcoRI/XhoI sites using the following 5' adaptor: GGCACGAG(G). Size-selected >500bp for average insert size 1.8kb. Library constructed by Ling Hong in the laboratory of Gerald M. Rubin (University of California, Berkeley) using ZAP-cDNA synthesis kit (Stratagene) and Superscript II RT (Life Technologies). Note: this is a NIH_MGC Library."
BASE COUNT      191 a      320 c      284 g      218 t      25 others
ORIGIN
Query Match      46.1%; Score 643.6; DB 14; Length 1038;
Best Local Similarity 88.6%; Pred. No. 5e-99;
Matches 748; Conservative 0; Mismatches 89; Indels 7; Gaps 5;

420 GTTACCTCTGACCTGCCCCCAGGTGCGACGCTGCTTTTTCAGACTGATGAGC 479
72 GTACACCAACGACCACTGAGACGTCGGAGCTGCTTTTTCAGACTGATGAGC 131

```

QY 480 CAAAGTGTCCCTGATCCCAACAGACCATATGTGAAGGCTCTGCTGACTATCTGA 539  
 DB 132 CAAAGTGTCCCTGATCCCAACAGACCATATGTGAAGGCTCTGCTGACTATCTGA 191  
 QY 540 GGGCTGGCTGACCACTGATCTTCTGACAGCTGGCTGCTGTGAGGAGTACT 599  
 DB 192 GGGCTGGCTGACCACTGATCTTCTGACAGCTGGCTGCTGTGAGGAGTACT 251  
 QY 600 TGCACTGGAGCACTGATCTGACCTGAGGAAACCCCTGACAGACCAAGCTAACATCCAGAC 659  
 DB 252 TGCACTGGAGCACTGATCTGACCTGAGGAAACCCCTGACAGACCAAGCTAACATCCAGAC 311  
 QY 660 AGACAGATGTGACCAAGACCAAGCTGATCTGACCAAGCTGATCTGACCAAGCTGATCTGACCA 719  
 DB 312 AGACAGATGTGACCAAGACCAAGCTGATCTGACCAAGCTGATCTGACCAAGCTGATCTGACCA 371  
 QY 720 GCTTACGATGTGACCAAGCTGATCTGACCAAGCTGATCTGACCAAGCTGATCTGACCAAGCT 778  
 DB 372 GCTTACGATGTGACCAAGCTGATCTGACCAAGCTGATCTGACCAAGCTGATCTGACCAAGCT 431  
 QY 779 CAACCTGTGGGCTGTGATGACCAAGCTGATCTGACCAAGCTGATCTGACCAAGCTGATCTGACCA 837  
 DB 432 CAACCTGTGGGCTGTGATGACCAAGCTGATCTGACCAAGCTGATCTGACCAAGCTGATCTGACCA 491  
 QY 838 CCTGGAGCGGTTCCGTGGGAGCCCATCACTGTGTTCAATGATGATGATGATGATGATGATGATGAT 897  
 DB 492 CCTGGAGCGGTTCCGTGGGAGCCCATCACTGTGTTCAATGATGATGATGATGATGATGATGATGAT 551  
 QY 898 AGCCCTG 957  
 DB 552 AGCCCTG 611  
 QY 958 CCATCGGAGGATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1017  
 DB 612 CCATCGGAGGATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 671  
 QY 1018 GATGTGGGATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1077  
 DB 672 GATGTGGGATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 731  
 QY 1078 CAGGCCAGCCAGGGGCTGAGGGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 1134  
 DB 732 CAGGCCAGCCAGGGGCTGAGGGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 791  
 QY 1135 GGTGAGG 1192  
 DB 792 GGTGAGG 851  
 QY 1193 TTCTGCTCATTTGCTTTCATGAAAGCTCAAGAGCAAGCAAGCAAGCAAGCAAGCAAGCAAGCAAG 1252  
 DB 852 CTTTTCCCTCCCTCTTTTCTTTTAAAGGCGCGCCGCGCCGCGCCGCGCCGCGCCGCGCTT 911  
 QY 1253 CTTT 1256  
 DB 912 CCCC 915

RESULT 3  
 BM671384/c 773 bp mRNA linear EST 27-FEB-2002  
 LOCUS BM671384  
 DEFINITION UI-E-CK1-afk-m-06-0-UI.82 UI-E-CK1 Homo sapiens cDNA clone  
 ACCESSION BM671384  
 VERSION BM671384.1 GI:18981282  
 KEYWORDS EST.  
 SOURCE human.  
 ORGANISM Homo sapiens  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.  
 REFERENCE 1 (bases 1 to 773)  
 AUTHORS Bonaldo, M.F., Lennon, G. and Soares, M.B.  
 TITLE Normalization and subtraction: two approaches to facilitate gene  
 discovery

JOURNAL  
MEDLINE  
COMMENT

Genome Res. 6 (9), 791-806 (1996)  
 97044477  
 Contact: Soares, MB  
 Program for Rat Gene Discovery and Mapping  
 University of Iowa  
 451 Eckstein Medical Research Building Iowa City, IA 52242, USA  
 Tel: 319 335 8250  
 Fax: 319 335 9565  
 Email: msoares@blue.weeg.uiowa.edu  
 Tissue Procurement: Dr. Gregg Hageman  
 CDNA library preparation: Dr. M. Bento Soares, University of Iowa  
 CDNA library Arrayed by: Dr. M. Bento Soares, University of Iowa  
 DNA Sequencing by: Dr. M. Bento Soares, University of Iowa  
 Clone Distribution: Researchers may obtain clones from Research  
 Genetics (www.resgen.com).  
 The following repetitive elements were found in this cDNA  
 sequence: 1-59, >(TTAA)n#simple\_repeat (matched complement)  
 Seq primer: M13 forward  
 POLYA=yes.

## FEATURES

## source

Location/Qualifiers  
 1..773  
 /organism="Homo sapiens"  
 /db\_xref="taxon:9606"  
 /clone="UI-E-CK1-afk-m-06-0-UI"  
 /clone\_1b="UI-E-CK1"  
 /tissue\_type="Retina Foveal and Macular"  
 /dev\_stage="adult"  
 /lab\_host="DH10B (Life Technologies) (T1 phage resistant)"  
 /note="Organ: eye; Vector: p773-Pac (Pharmacia) with a  
 modified polylinker; Site 1: EcoR I; Site 2: Not I;  
 UI-E-CK1 is a normalized cDNA library containing the  
 following tissue(s): Retina Foveal and Macular. The  
 library was constructed according to Bonaldo, Lennon and  
 Soares, Genome Research, 6:791-806, 1996. First strand  
 cDNA synthesis was primed with an oligo-dT primer  
 containing a Not I site. Double stranded cDNA was ligated  
 to an EcoR I adaptor, digested with Not I, and cloned  
 directionally into p773-Pac vector. The oligonucleotide  
 used to prime the synthesis of first-strand cDNA contains  
 a library tag sequence that is located between the Not I  
 site and the (dT)18 tail. The sequence tag for this  
 library is GTCC. This library was created for the program,  
 Gene Discovery in the Visual System, supported by National  
 Eye Institute (NEI).  
 TAG LIB=UI-E-CK1  
 TAG\_TISSUE=Foveal and Macular Retina  
 TAG\_SEQ=GTCC"

BASE COUNT 181 a 221 c 180 g 189 t 2 others  
 ORIGIN

Query Match 44.8%; Score 626.4; DB 14; Length 773;  
 Best Local Similarity 98.2%; Pred. No. 4,6e-96;  
 Matches 644; Conservative 0; Mismatches 11; Indels 1; Gaps 1;  
 QY 720 GCTTACGATGTGACCAAGCTGATCTGACCAAGCTGATCTGACCAAGCTGATCTGACCAAGCTGATCTGACCA 779  
 DB 656 GCTTACGATGTGACCAAGCTGATCTGACCAAGCTGATCTGACCAAGCTGATCTGACCAAGCTGATCTGACCA 597  
 QY 780 -AACCTGTGGGCTGTGACCAAGCTGATCTGACCAAGCTGATCTGACCAAGCTGATCTGACCAAGCTGATCTGACCA 838  
 DB 596 NAACCTGTGGGCTGTGACCAAGCTGATCTGACCAAGCTGATCTGACCAAGCTGATCTGACCAAGCTGATCTGACCA 537  
 QY 839 CTGGAGCGGTTCCGTGGGAGCCCATCACTGTGTTCAATGATGATGATGATGATGATGATGATGATGAT 898  
 DB 536 CTGGAGCGGTTCCGTGGGAGCCCATCACTGTGTTCAATGATGATGATGATGATGATGATGATGATGAT 477  
 QY 899 GCCCTGTGCTG 958  
 DB 476 GCCCTGTGCTG 417  
 QY 959 CATGTGAGTGTCTTCTGAGCTTGTGAGCAGAGAGCTTGGCGGGAGTGTCCAG 1018  
 DB 416 CATGTGAGTGTCTTCTGAGCTTGTGAGCAGAGAGCTTGGCGGGAGTGTCCAG 357



University of Iowa  
451 Eckstein Medical Research Building Iowa City, IA 52242, USA  
Tel.: 319 335 9250  
Fax: 319 335 9365  
Email: msouares@blue.weeg.uiowa.edu  
Tissue Procurement: Dr. Gregg Hageman  
CDNA Library preparation: Dr. M. Bento Soares, University of Iowa  
CDNA Library Arrayed by: Dr. M. Bento Soares, University of Iowa  
DNA Sequencing by: Dr. M. Bento Soares, University of Iowa  
Clone Distribution: Researchers may obtain clones from Research  
Genetics ([www.researchgen.com](http://www.researchgen.com)).  
The following repetitive elements were found in this CDNA  
sequence: 1-59, >(TAAA)n\$imple\_repeat (matched complement)  
Seq primer: M13 Forward  
POLYA=yes.

FEATURES	Location/Qualifiers
source	1. .645

```

source
1. .645
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="UT-E-DX1-agn-1-21-0-UT"
/clone_lib="UT-E-DX1"
/tissue_type="fetal eyes"
/dev_stage="fetal"
/lab_host="DH10B (Life Technologies) (T1 phage resistant)"
/notes="Organ: eye; Vector: pT73-pac (Pharmacia) with a
modified polylinker; Site_1: EcoR I; Site_2: Not I;
UT-E-DX1 is a normalized cDNA library containing the
following tissue(s): fetal eyes. The library was
constructed according to Bonaldi, Lemon and Soares,
Genome Research, 6:791-806, 1996. First strand cDNA
synthesis was primed with an oligo-dT primer containing a
Not I site. Double stranded cDNA was ligated to an EcoR I
adaptor, digested with Not I, and cloned directionally
into pT73-pac vector. The oligonucleotide used to prime
the synthesis of first-strand cDNA contains a library tag
sequence that is located between the Not I site and the
(dT)18 tail. The sequence tag for this library is
AGAAATCAGA. This library was created for the program, Gene
Discovery in the Visual System, supported by National Eye
Institute (NEI).
TAG_LIB=UT-E-DX1
TAG_TISSUE=human fetal eyes
TAG_SEQ=AGAAATCAGA"
152 a 188 c 155 g 150 t

```

Query Match	43.1%	Score 602.8	DB 13	Length 645
Best Local Similarity	97.8%	Pred. No. 4,7e-92		
Matches 632	Conservative 0	Mismatches 12	Indels 2	Gaps 2
QY	730	GGGAGCTGCTGCTCCTTAAGTCCAGGAATCATGGGGGTATGACTGCCTCTCCAA-CCCTGTG	788	
Db	645	GGGAGCTGTGGTTCCTAAGTCCAGGAATCATGGGGGTATGACTGCCTCTCCAAACCCCTGTG	586	
QY	789	GGCTGTAGGCAAGCTCAGGCTAAGTCTCCCACTGGGGGCTGTGCCCTCTCTGGGAGCGT	848	
Db	585	GGCTGTAGGCAAGCTCAGGCTAAGTCTCCCACTGGGGGCTGTGCCCTCTCTGGGAGCGT	526	
QY	849	TCCGAGGGACGCCCATCATCTGTGTTCAATAGTGTGAATGTAATGACTTAAAGCCCTGCTG	908	
Db	525	TCCGTGGGACGCCCATCATCTGTGTTCAATAGTGTGAATGTAATGACTTAAAGCCCTGCTG	466	
QY	909	CTGCTGTGCACATGCCACACAGAGCGGTGGGGGCTGCTGGGGGACATTCATCGTGAG	968	
Db	465	CTGCTGTGCACATGCCACACAGAGCGGTGGGGGCTGTGCT-GGGAACAATCATGCTGAG	407	
QY	969	TGTTCTCTCAGCTTAGTGTGTGGACAGGAGACTTGCGGGGGATGCTCCAGGATGTGGGTG	1028	
Db	406	TGTTCTCTCAGCTTAGTGTGTGGACAGGAGACTTGCGGGGGATGCTCCAGGATGTGGGTG	347	
QY	1029	ATTCTGTACCTGGGGAGGCTATCTCTGACCTCCGACAGGGGAGACATCCACAGGACAGCC	1088	
Db	346	ATTCTGTACCTGGGGAGGCTATCTCTGACCTCCGACAGGGGAGACATCCACAGGACAGCC	287	

[illegible]

## RESULT 6

LOCUS	BM668538	610 bp	mRNA	linear	EST 27-FEB-2002
DEFINITION	U1-E-CX1-afm-C-23-0-U1.s2 U1-E-CX1 Homo sapiens CDNA clone. U1-E-CX1-afm-C-23-0-U1 3', mRNA sequence.				
ACCESSION	BM668538				
VERSION	BM668538.1				
KEYWORDS	EST.				
SOURCE	human.				
ORGANISM	Homo sapiens				

## ORGANISM

REFERENCE	AUTHORS	TITLE	JOURNAL	MEDLINE	COMMENT
1	(bases 1 to 610)	Mammalia: Eutheria; Primates; Carnivora; Homnidae; Homo.			
	Bonaldi, M. F., Lennon, G. and Soares, M. B.	Normalization and subtraction: two approaches to facilitate gene discovery	Genome Res. 6 (9), 791-806 (1996)		Contact: Soares, MB

COMMENT

University of Iowa  
451 Eckstein Medical Research Building Iowa City, IA 52242, USA  
Tel.: 319 335 8250  
Fax: 319 335 9365  
Email: mscares@blue.weeg.uiowa.edu  
Tissue Procurement: Dr. Gregg Hegeman  
cDNA library preparation: Dr. M. Bento Soares, University of Iowa  
cDNA library Arrayed by: Dr. M. Bento Soares, University of Iowa  
DNA Sequencing by: Dr. M. Bento Soares, University of Iowa  
Clone Distribution by: Researchers may obtain clones from Research  
Genetics ([www.resgen.com](http://www.resgen.com)).  
The following repetitive elements were found in this cDNA  
sequence: 32-80, >AT-richLow\_complexity  
Seq primer: M13 Forward  
POLYA=yes.

## FEATURES

```

/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="UI-E-CX1-afm-c-23-0-UI"
/clone_lib="UI-E-CX1"
/tissue_type="Retina Foveal and Macular"
/dev_stage="adult"
/lab_host="DH10B (Life Technologies) (T1 phage resistant)"
/note="Organ: eye; Vector: pRTT3-Pac (Pharmacia) with a
modified polylinker; Site_1: EcoR I; Site_2: Not I;
UI-E-CX1 is a normalized cDNA library containing the
following tissue(s): Retina Foveal and Macular. The
library was constructed according to Bonaldo, Lennon and
Scores, Genome Research, 6:791-806, 1996. First strand
cDNA synthesis was primed with an oligo-dT primer

```

containing a Not I site. Double stranded cDNA was ligated to an EcoR I adaptor, digested with Not I, and cloned directionally into pT73-Pac Vector. The oligonucleotide used to prime the synthesis of first-strand cDNA contains a library tag sequence that is located between the Not I site and the (dr)18 tail. The sequence tag for this library is GTCC. This library was created for the program, Gene Discovery in the Visual System, supported by National Eye Institute (NEI).

Eye Institute (NEI).  
TAG: LIB=U1-E-CX1  
TAG: TISSUE=Foveal and Macular Retina  
TAG: SEQ=GTCC

BASE COUNT 145 a 147 g 143 t 1 others  
ORIGIN

Query Match 42.4%; Score 593; DB 13; Length 610;  
Best Local Similarity 98.2%; Pred. No. 2.2e-90;  
Matches 599; Conservative 0; Mismatches 11; Indels 0; Gaps 0;

765 TATGACTGCTCTCCACCCCTGTTGGGCTGTAGACAGCTGAGCTAGTCTCCCACTGGG 824  
610 TATGACTGCTCTCCACCCCTGTTGGGCTGTAGACAGCTGAGCTAGTCTCCCACTGGG 551  
825 GGGTGGCCCTCCCTGCGAGCGTTCCGTGGGAGCCCACTGTTGTTCAATAGTGTG 884  
550 GGGTGGCCCTCCCTGCGAGCGTTCCGTGGGAGCCCACTGTTGTTCAATAGTGTG 491  
885 AGAATGATGCTAAAGCCCTGCTGCTGCTGCAATCCCAAGCGGGGTGGGCT 944  
490 AGAATGATGCTAAAGCCCTGCTGCTGCTGCAATCCCAAGCGGGGTGGGCT 431  
945 GCGTGGGACAAATCCATGCTGAGTGTCTCTGAGTGTAGGCTGAGCAGAGACTGGC 1004  
430 GCGTGGGACAAATCCATGCTGAGTGTCTCTGAGTGTAGGCTGAGCAGAGACTGGC 371  
1005 GGGGATGCTCCAGATGTGGGTATCTGTACCTGGGAGGCTATCTGACCTCCGA 1064  
370 GGGGATGCTCCAGATGTGGGTATCTGTACCTGGGAGGCTATCTGACCTCCGA 311  
1065 CAGGGGACACTCCCAAGGCGAGCCCAAGGGGTGAGGGGAGGAGTCACTGAGCTAGG 1124  
310 CAGGGGACACTCCCAAGGCGAGCCCAAGGGGTGAGGGGAGGAGTCACTGAGCTAGG 251  
1125 CCAAGACTGGGGTTCAGGAGCAGGTGTGTTGAGCAGAGACTGGGGGCGGGGTGGG 1184  
250 CCAAGACTGGGGTTCAGGAGCAGGTGTGTTGAGCAGAGACTGGGGGCGGGGTGGG 191  
1185 CCGGGCTCTTCTGCTGCTATTTGCTTCAATGAAGCTCAAGAGCCCAAGCGCTT 1244  
190 CCGGGCTCTTCTGCTGCTATTTGCTTCAATGAAGCTCAAGAGCCCAAGCGCTT 131  
1245 TCCCTCTCTGAGTTGATATCCAGAACTTTTGTACTTGTGTGTTAAATGTT 1304  
130 TCCCTCTCTGAGTTGATATCCAGAACTTTTGTACTTGTGTGTTAAATGTT 71  
1305 TATTTTGTAAATAAATAAATAAATTAGTTAATAAATGATGTTTCAAGCAAACTCTT 1364  
70 TATTTTGTAAATAAATAAATAAATTAGTTAATAAATGATGTTTCAAGCAAACTCTT 11  
1365 CCTTAAAAA 1374  
10 AAAAAAAAAA 1

RESULT 7  
AM409897/c  
LOCUS  
DEFINITION 678 bp mRNA linear EST 29-JUN-2000  
fh030806.y1 NIH\_MGC\_17 Homo sapiens cDNA clone IMAGE:2960938 3',  
mRNA sequence.  
ACCESSION AM409897  
VERSION AM409897  
KEYWORDS EST.  
SOURCE human.  
GI:6935438

ORGANISM Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.  
1 (bases 1 to 678)  
REFERENCE NIH-MGC http://mgc.nci.nih.gov/  
AUTHORS National Institutes of Health, Mammalian Gene Collection (MGC)  
TITLE Unpublished (1999)  
JOURNAL Contact: Robert Strausberg, Ph.D.  
COMMENT Email: cgabbs-remail.nih.gov  
Tissue Procurement: ATCC  
CDNA Library Preparation: Ling Hong/Rubin Laboratory  
CDNA Library Arrayed by: The I.M.A.G.E. Consortium (ILNL)  
DNA Sequencing by: National Institutes of Health Intramural  
Sequencing Center (NISC)  
Clone distribution: MGC clone distribution information can be  
found through the I.M.A.G.E. Consortium/ILNL at:  
www.bio.lnsl.gov/bdrip/image/image.html  
Plate: LLCM55 row: B column: 11  
Seq primer: M13RP1 reverse primer (ABI).

FEATURES  
source  
1..678  
/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
/clone="IMAGE:2960938"  
/clone\_lib="NIH\_MGC\_17"  
/clone\_type="rhabdomyosarcoma"  
/lab\_host="DH10B (phage-resistant)"  
/note="Organ: muscle; Vector: pORF7; Site 1: EcoRI;  
Site 2: XhoI; cDNA made by oligo-dT priming.  
Directionally cloned into EcoRI/XhoI sites using the  
following 5' adaptor: GGCACGAC(G). Size-selected 500bp  
for average insert size 1.8kb. Library constructed by  
Ling Hong in the laboratory of Gerald M. Rubin (University  
of California, Berkeley) using ZAP-cDNA synthesis kit  
(Stratagene) and Superscript II RT (Life Technologies)."

BASE COUNT 161 a 190 c 159 g 168 t  
ORIGIN

Query Match 42.1%; Score 588.4; DB 10; Length 678;  
Best Local Similarity 92.8%; Pred. No. 1.2e-89;  
Matches 648; Conservative 0; Mismatches 1; Indels 49; Gaps 1;

645 GCTAATCCCGACAGACAGATGTGACAGAGCAAAAGTGTCAATAGCCAAATGTTAA 704  
678 GCTAATCCCGACAGACAGATGTGACAGAGCAAAAGTGTCAATAGCCAAATGTTAA 619  
705 AATGTAGTTTACACGCTAGTATGAGTGTGCTGCTAGTCCAGAAATCATAGGGGG 764  
618 AATGTAGTTTACACGCTAGTATGAGTGTGCTGCTAGTCCAGAAATCATAGGGGG 559  
765 TATGACTGCTCTCCAAACCTGTGGGCTGTAGCAAGCTGAGGTAATCTCCCACTGGG 824  
558 TATGACTGCTCTCCAAACCTGTGGGCTGTAGCAAGCTGAGGTAATCTCCCACTGGG 499  
825 GCGTGGCCCTCCCTGCGAGCGTTCCGTGGGAGCCCACTGTTGTTCAATAGTGTG 884  
498 GCGTGGCCCTCCCTGCGAGCGTTCCGTGGGAGCCCACTGTTGTTCAATAGTGTG 439  
885 AGAATGATGCTAAAGCCCTGCTGCTGCTGCAATGCCAAGCGGGGTGGGCT 944  
438 AGAATGATGCTAAAGCCCTGCTGCTGCTGCAATGCCAAGCGGGGTGGGCT 428  
945 GCGTGGGACAAATCCATGCTGAGTGTCTCTGAGTGTAGGCTGAGCAGAGACTGGC 1004  
427 GCGTGGGACAAATCCATGCTGAGTGTCTCTGAGTGTAGGCTGAGCAGAGACTGGC 368  
1005 GGGGATGCTCCAGAGTGGGTATCTGTAAGTGGGAGGCTATCTGTAAGTGGGAG 1064  
367 GGGGATGCTCCAGAGTGGGTATCTGTAAGTGGGAGGCTATCTGTAAGTGGGAG 308  
1065 CAGGGGACACTCCCAAGGCGAGCCCAAGGGGTGAGGGGAGAGTGCACACCTGACATGAG 1124  
307 CAGGGGACACTCCCAAGGCGAGCCCAAGGGGTGAGGGGAGAGTGCACACCTGACATGAG 248



QY 1125 CCAAGACTGGGCTCAGGAGCAGGTGTGTTTGAAGCCAGCCTGGGGCGGGGTGGGCG 1184  
DB 247 CCAAGACTGGGCTCAGGAGCAGGTGTGTTTGAAGCCAGCCTGGGGCGGGGTGGGCG 188  
QY 1185 CGGGGCTTCTTCCCTCATTTGCTTTCATGAAAGCTCAAGAGCCAAACAGGCTT 1244  
DB 187 CGGGGCTTCTTCCCTCATTTGCTTTCATGAAAGCTCAAGAGCCAAACAGGCTT 128  
QY 1245 TCCCTCTCTCAGATTGATATCCAGATCTTTTACTCTCTGTGTGTTAAATGTT 1304  
DB 127 TCCCTCTCTCAGATTGATATCCAGATCTTTTACTCTCTGTGTGTTAAATGTT 68  
QY 1305 TATTTTGTAAAAATAAATAAATAAATTAGTTAATAAAA 1342  
DB 67 TATTTTGTAAAAATAAATAAATAAATTAGTTAATAAAA 30

RESULT 8  
BG827019 776 bp mRNA linear EST 22-MAY-2001  
LOCUS 602749053F1 NIH\_MGC\_17 Homo sapiens cDNA clone IMAGE:4901869 5',  
DEFINITION mRNA sequence.  
ACCESSION BG827019 GI:14174606  
VERSION BG827019.1  
KEYWORDS EST.  
SOURCE human.  
ORGANISM Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Hominiidae; Homo.  
REFERENCE 1. (bases 1 to 776)  
NIH-MGC http://mgs.nci.nih.gov/.  
National Institutes of Health, Mammalian Gene Collection (MGC)  
Unpublished (1999)  
COMMENT Contact: Robert Strausberg, Ph.D.  
Email: cgabbs-remail.nih.gov  
Tissue Procurement: ATCC  
CDNA Library Preparation: Ling Hong/Rubin Laboratory  
CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LNL)  
DNA Sequencing by: Incyte Genomics, Inc.  
Clone distribution: MGC clone distribution information can be  
found through the I.M.A.G.E. Consortium/LNL at:  
http://image.llnl.gov  
Plate: LNCM1797 row: j column: 14  
High quality sequence stop: 771.

FEATURES  
source Location/Qualifiers  
1..776  
/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
/clone="IMAGE:4901869"  
/clone\_1ib="NIH MGC 17"  
/tissue\_type="fibrocytoma"  
/lab\_host="DH10B (phage-resistant)"  
/note="Organ: muscle; Vector: pOTB7; Site: 1: EcoRI;  
Site: 2: XhoI; cDNA made by oligo-dT priming.  
Directionally cloned into EcoRI/XhoI sites using the  
following 5' adaptor: GCGACGAG(G). Size-selected >500bp  
for average insert size 1.8kb. Library constructed by  
Ling Hong in the laboratory of Gerald M. Rubin (University  
of California, Berkeley) using ZAP-cDNA synthesis kit  
(Stratagene) and Superscript II RT (Life Technologies)."  
BASE COUNT 158 a 228 c 168 t  
ORIGIN

Query Match 41.1%; Score 574.6; DB 12; Length 776;  
Best Local Similarity 97.7%; Pred. No. 2.4e-87;  
Matches 583; Conservative 0; Mismatches 14; Indels 0; Gaps 0;

QY 420 GTTACTCTCAGCTTGGCCCGGAGCTGCTGTTTCAAGACTGGATGGAG 479  
DB 180 GTTACTCTCAGCTTGGCCCGGAGCTGCTGTTTCAAGACTGGATGGAG 239  
QY 480 CAAGTGTCCCTGATCCCAAGACCAATATGTGAAGGCTCTGTGCTGACTATCTGA 539

DB 240 CAAGTGTCCCTGATCCCAAGACCAATATGTGAAGGCTCTGTGCTGACTATCTGA 299  
QY 540 GGGCTGGGTCACACACTACTACTCTCAGCAGCTGGGGCTGTCTGTGAGGGAGTACT 599  
DB 300 GGGCTGGGTCACACACTACTACTCTCAGCAGCTGGGGCTGTCTGTGAGGGAGTACT 359  
QY 600 TGCACATGGAGACACTGCATGTCACTGGGAACCCCTGCAGACAAAGCTAATCCAGAC 659  
DB 360 TGCACATGGAGACACTGCATGTCACTGGGAACCCCTGCAGACAAAGCTAATCCAGAC 419  
QY 660 AGACAGATGTGACACAGACAAAGCTGCAATATGCAAAATGTTAAATGTAGTTTACA 719  
DB 420 AGACAGATGTGACACAGACAAAGCTGCAATATGCAAAATGTTAAATGTAGTTTACA 479  
QY 720 GCTTACTATGGGACCTGCTGCTCTCTAGTCCAGAGATTCATGGGGGTATGACTCTCTCC 779  
DB 480 GCTTACTATGGGACCTGCTGCTCTCTAGTCCAGAGATTCATGGGGGTATGACTCTCTCC 539  
QY 780 AACCTGTGGGCTGTAGGACAGCTCAGGCTAGTCTCCCACTGGGGGTGTGCTCCCTCCG 839  
DB 540 AACCTGTGGGCTGTAGGACAGCTCAGGCTAGTCTCCCACTGGGGGTGTGCTCCCTCCG 599  
QY 840 TGGAGCGTTCCGTGGGACAGCCCATCACTGTGTTCAATAGTGAATGTAGTTAAG 899  
DB 600 TGGAGCGTTCCGTGGGACAGCCCATCACTGTGTTCAATAGTGAATGTAGTTAAG 659  
QY 900 CCCCTGCTCTCTGTGTGACATGCGACAGCAGCGGTGGGGCTGCTGGGACATATC 959  
DB 660 CCCCTGCTCTCTGTGTGACATGCGACAGCAGCGGTGGGGCTGCTGGGACATATC 719  
QY 960 ATCTGTGAGTGTCTCTCAGCTTAGTGTGACAGAGACTTGGCGGGGATGCTCC 1016  
DB 720 ATCTGTGAGTGTCTCTCAGCTTAGTGTGACAGAGACTTGGCGGGGATGCTCC 776

RESULT 9  
BE250262 770 bp mRNA linear EST 13-JUL-2000  
LOCUS 600943273F1 NIH\_MGC\_17 Homo sapiens cDNA clone IMAGE:2959770 5',  
DEFINITION mRNA sequence.  
ACCESSION BE250262  
VERSION BE250262.1 GI:9120370  
KEYWORDS EST.  
SOURCE human.  
ORGANISM Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Hominiidae; Homo.  
REFERENCE 1. (bases 1 to 770)  
NIH-MGC http://mgs.nci.nih.gov/.  
National Institutes of Health, Mammalian Gene Collection (MGC)  
Unpublished (1999)  
COMMENT Contact: Robert Strausberg, Ph.D.  
Email: cgabbs-remail.nih.gov  
Tissue Procurement: ATCC  
CDNA Library Preparation: Ling Hong/Rubin Laboratory  
CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LNL)  
DNA Sequencing by: Incyte Genomics, Inc.  
Clone distribution: MGC clone distribution information can be  
found through the I.M.A.G.E. Consortium/LNL at: image.llnl.gov  
Plate: LNCM52 row: a column: 19  
High quality sequence stop: 730.

FEATURES  
source Location/Qualifiers  
1..770  
/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
/clone="IMAGE:2959770"  
/clone\_1ib="NIH MGC 17"  
/tissue\_type="fibrocytoma"  
/lab\_host="DH10B (phage-resistant)"  
/note="Organ: muscle; Vector: pOTB7; Site: 1: EcoRI;  
Site: 2: XhoI; cDNA made by oligo-dT priming.  
Directionally cloned into EcoRI/XhoI sites using the



following 5' adaptor: GGACGAG(G). Size-selected >500bp for average insert size 1.8kb. Library constructed by Ling Hong in the laboratory of Gerald M. Rubin (University of California, Berkeley) using ZAP-cDNA synthesis kit (Stratagene) and Superscript II RT (Life Technologies)."

BASE COUNT 157 a 219 c 226 g 168 t

Query Match 40.3%; Score 562.6; DB 10; Length 770;  
Best Local Similarity 97.4%; Pred. No. 2.5e-85;  
Matches 593; Conservative 0; Mismatches 14; Indels 2; Gaps 2;

420 GTTACCTCAGCTGCGCCCGAGGTGGAGCTGCTCTTTTCAAGACTGATGAGC 479  
163 GTTACCAACCAACCACTGAGAGGTGGAGCTGCTCTTTTCAAGACTGATGAGC 222  
480 CAAGTGGTCCCTGATCCCAACCAACCACTGATGAGAGCTGCTGATCTATCTGA 539  
223 CAAGTGGTCCCTGATCCCAACCAACCACTGATGAGAGCTGCTGATCTATCTGA 282  
540 GGGCTGGCTGACCACTGATCTCTCTGACAGCTGGGCTTGGCTGGAGAGTGACT 539  
283 GGGCTGGCTGACCACTGATCTCTCTGACAGCTGGGCTTGGCTGGAGAGTGACT 342  
600 TGACCTGGAGCACTGATCTCTCTGAGAACCTGAGAACCAAGCTATCCAGAC 659  
343 TGACCTGGAGCACTGATCTCTCTGAGAACCTGAGAACCAAGCTATCCAGAC 402  
660 AGACAGATGTACCAAGCAACAGTGCATATATGCTTAAATGTGAGTTTACCA 719  
403 AGACAGATGTACCAAGCAACAGTGCATATATGCTTAAATGTGAGTTTACCA 462  
720 GCTTACGTATGAGAGCTGCTGCTCTCTGATCCAGAACTATGAGGCTCTCTCC 779  
463 GCTTACGTATGAGAGCTGCTGCTCTCTGATCCAGAACTATGAGGCTCTCTCC 522  
780 AACCTGTGGAGCTGTAAGCAAGCTCAGGCTAGTCTCCCACTGGGGCTGTGCCCTCC 839  
523 AACCTGTGGAGCTGTAAGCAAGCTCAGGCTAGTCTCCCACTGGGGCTGTGCCCTCC 582  
840 TGAGACGCTTCCGTGGGAGCCCACTGCTGTCTTCAATAGTGAATGTGACTTAAG 899  
583 TGAGACGCTTCCGTGGGAGCCCACTGCTGTCTTCAATAGTGAATGTGACTTAAG 642  
900 CCCCTGTGCTGCTGCTGCAATGCAACAGAGCGGTGGGCTGTGGTGGGCAATCC 959  
643 CCCCTGTGCTGCTGCTGCAATGCAACAGAGCGGTGGGCTGTGGTGGGCAATCC 701  
960 ATCGTGAAGTGTCTCTCAGCTTGTGAGAGCACTGGCGGGAGTGTCTCAGG 1019  
702 ATCGTGAAGTGTCTCTCAGCTTGTGAGAGCACTGGCGGGAGTGTCTCAGG 760  
1020 ATGTGGGTG 1028  
761 ATGTGGGTG 769

RESULT 10  
AM516922/c  
LOCUS  
DEFINITION  
3', mRNA sequence.  
565 bp mRNA linear EST 03-MAR-2000  
X888909.x1 Soares\_NHCE\_cervix Homo sapiens cDNA clone IMAGE:2747488

AM516922  
AM516922.1 GI:7154931

ACCESSION  
VERSION  
KEYWORDS  
SOURCE  
ORGANISM  
human.  
Homo sapiens  
Eukaryota; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

REFERENCE  
1 (bases 1 to 565)  
NCI-CCAP <http://www.ncbi.nlm.nih.gov/ncicgap>.  
National Cancer Institute, Cancer Genome Anatomy Project (CGAP),  
Tumor Gene Index

JOURNAL  
COMMENT  
Unpublished (1997)  
Contact: Robert Strausberg, Ph.D.  
Email: [cgapsb-remail.nih.gov](mailto:cgapsb-remail.nih.gov)  
This clone is available royalty-free through LNL; contact the  
IMAGE Consortium ([info@image.llnl.gov](mailto:info@image.llnl.gov)) for further information.  
Seq primer: -40UP from Gibco  
High quality sequence stop: 450.

FEATURES  
source  
1..565  
/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
/clone="IMAGE:2747488"  
/lab\_host="PH10B (phage-resistant)"  
/note="Organ: cervix; Vector: pT73D-Pac (Pharmacia) with  
a modified polylinker; Site 1: Not I; Site 2: Eco RI;  
strand cDNA was primed with a Not I - oligo(dT) primer [5'  
TGTTACATCTGAAGTGGAGCGGCGCGGCTTTTCTTTTCTTTTCTTTTCTTTT  
T 3']; double-stranded cDNA was ligated to Eco RI  
adaptors (Pharmacia), digested with Not I and cloned into  
the Not I and Eco RI sites of the modified pT73 vector.  
Library is normalized; constructed by Bento Soares and  
M.Fatima Bonaldo."

BASE COUNT 138 a 169 c 136 g 121 t 1 others

Query Match 40.3%; Score 562.4; DB 10; Length 565;  
Best Local Similarity 99.6%; Pred. No. 3.2e-85;  
Matches 563; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

791 CTGTAAGCAAGCTCAGGCTAGTCTCCCACTGGGGCTGTGCTCTCCCTGAGAGCTTC 850  
565 CTGTAAGCAAGCTCAGGCTAGTCTCCCACTGGGGCTGTGCTCTCCCTGAGAGCTTC 506  
851 CGTGGAGCCCACTCAGTCTGTTCAATATGTGTGAATATGTGCTAAAGCCCTGTGCT 910  
505 CGTGGAGCCCACTCAGTCTGTTCAATATGTGTGAATATGTGCTAAAGCCCTGTGCT 446  
911 GGTGCTGCATATGCCACAGCAGCGGTGGGCTGTGCTGGGAGCAATCATGTGAGAGT 970  
445 GGTGCTGCATATGCCACAGCAGCGGTGGGCTGTGCTGGGAGCAATCATGTGAGAGT 386  
971 TTCTCAGCTTAAAGTCTGAGACAGAGCTTGGCGGGAGTCTCCAGATGTGGTAT 1030  
385 TTCTCAGCTTAAAGTCTGAGACAGAGCTTGGCGGGAGTCTCCAGATGTGGTAT 326  
1031 TCTGTACTGGGAGGCTTCTCTGACCTCCGACAGGGGACACTCCAGGCCAGCCAG 1090  
325 TCTGTACTGGGAGGCTTCTCTGACCTCCGACAGGGGACACTCCAGGCCAGCCAG 266  
1091 GGGTCAAGGGGAGAGGTGACACCTCAGCATGAGCCAAAGCTGGGGTCAAGGAGCAGTG 1150  
265 GGGTCAAGGGGAGAGGTGACACCTCAGCATGAGCCAAAGCTGGGGTCAAGGAGCAGTG 206  
1151 TGGTTTGAACCAAGACCTTGGGGGAGGGGCTGGGGCTTTTCTGCTCATTTGCTTT 1210  
205 TGGTTTGAACCAAGACCTTGGGGGAGGGGCTGGGGCTTTTCTGCTCATTTGCTTT 146  
1211 CAATGAAGCTTCAAGACAGCCAAAACAGGCTTTCCCTTCTGAGATTTGAATATCC 1270  
145 CAATGAAGCTTCAAGACAGCCAAAACAGGCTTTCCCTTCTGAGATTTGAATATCC 86  
1271 AGAATCTTTTGAATCTTGTGTTAAATGTTTATTTTGAATTAATTAATTAAT 1330  
85 AGAATCTTTTGAATCTTGTGTTAAATGTTTATTTTGAATTAATTAATTAAT 26  
1331 TACTTAATAAATGATGTTTCAAG 1355  
25 TACTTAATAAATGATGTTTCAAG 1

RESULT 11  
B0647652

LOCUS B0647652 935 bp mRNA linear EST 15-JUL-2002  
 DEFINITION AGENCOURT\_8288722 NIH\_MGC\_100 Homo sapiens cDNA clone IMAGE:6292938  
 5', mRNA sequence.  
 ACCESSION B0647652  
 VERSION B0647652.1 GI:21771824  
 KEYWORDS EST.  
 SOURCE human.  
 ORGANISM Homo sapiens  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.  
 REFERENCE 1 (bases 1 to 935)  
 NIH-MGC. <http://mgi.nci.nih.gov/>.  
 National Institutes of Health, Mammalian Gene Collection (MGC)  
 Unpublished (1999)  
 CONTACT: Robert Strausberg, Ph.D.  
 Email: cgaabs-remail.nih.gov  
 Tissue Procurement: CGAP (Stanford)  
 CDNA Library Preparation: Rubin Laboratory  
 CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)  
 DNA Sequencing by: Agencourt Bioscience Corporation  
 Clone distribution: MGC clone distribution information can be  
 found through the I.M.A.G.E. Consortium/LLNL at:  
<http://image.llnl.gov>  
 Plate: LHC2495 row: c column: 19  
 High quality sequence stop: 476.  
 Location/Qualifiers  
 1. 935  
 /organism="Homo sapiens"  
 /db\_xref="taxon:9606"  
 /clone\_1ib="NIH\_MGC\_100"  
 /issue\_type="hepatocellular carcinoma, cell line"  
 /lab\_host="DH10B (phage-resistant)"  
 /note="Organ: liver; Vector: pOTB7; Site 1: XhoI; Site 2:  
 EcoRI. cDNA made by oligo-dT priming. Directionally cloned  
 into EcoRI/XhoI sites using the following 5' adaptor:  
 GGACGAG(G). Size-selected >500bp for average insert size  
 1.8kb. Library constructed by Ling Hong in the laboratory  
 of Gerald M. Rubin (University of California, Berkeley)  
 using ZAP-cDNA synthesis kit (Stratagene) and Superscript  
 II RT (Life Technologies). Note: this is a NIH\_MGC  
 Library."  
 BASE COUNT 201 a 253 c 268 g 207 t 6 others  
 ORIGIN  
 Query Match 39.7%; Score 554; DB 14; Length 935;  
 Best Local Similarity 90.1%; Pred. No. 6.3e-84;  
 Matches 727; Conservative 0; Mismatches 63; Indels 17; Gaps 12;

QY 778 CCAACCTGTGGCTGTATAGCAAGCTACAGCTAGTCTCCCACTGGGGCTGTGCCCCCT 836  
 DB 432 CCAACCTGTGGCTGTATAGCAAGCTACAGCTAGTCTCCCACTGGGGCTGTGCCCCCT 491  
 QY 837 CCTGGGAGAGTTCCTGGGAGAGCCCAATCAGCTGTGTCAATAGTGTGAAGTAGCTA 896  
 DB 492 CCTGGGAGAGTTCCTGGGAGAGCCCAATCAGCTGTGTCAATAGTGTGAAGTAGCTA 551  
 QY 897 AAGCCCTGTGCTGTGTCGACATGACAC-AGCAGGCGGT-GGGGGCTGCTGGGGAGC 954  
 DB 552 AAGCCCTGTGCTGTGTCGACATGACACAGCAGAGGCGGTGGGGCTGCTGGGGAGC 611  
 QY 955 AATCCA-TGTGAGAGTTCCTGACAGCTTAGTGTGAGAGAGAC-TTGCGCGGGAGATG 1012  
 DB 612 AATCCATTGTGAGAGTTCCTGACAGCTTAGTGTGAGAGAGACCTTGCGCGGGAGATG 671  
 QY 1013 CTCGAGG-ATGTGGGTATTTCTGTACTTG-GGAGGCTATCTTACCTCCCGACAGGG 1069  
 DB 672 CTCGAGGAGTGTGGGTGTATTTCTGTACTTGAGGAGGCTATCTTACCTCCCGACAGG 731  
 QY 1070 GACACTCCAGGCGCAGCCCAAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 1126  
 DB 732 GACACTCCAGGCGCAGCCCAAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 791  
 QY 1127 -AAGACTGGGGTCAAGG--AGCAGGTGTGTTTGAAGCAGGACTGGGGCGGGGTGGG 1182  
 DB 792 CAAACTTGGGGTCAAGGAAACAAGTTGTGTTTGAAGCAGGAACTGGGGCGGGCGG 851  
 QY 1183 GCCGGGCGCTTGTGCTGCTATTTGCTT 1209  
 DB 852 GTGAGACCCCGGGCGCTTTTCTTCTT 878

RESULT 12  
 LOCUS B0692817 548 bp mRNA linear EST 28-FEB-2002  
 DEFINITION UI-E-CKI-afk-m-06-0-UI.r1 UI-E-CKI Homo sapiens cDNA clone  
 UI-E-CKI-afk-m-06-0-UI 5', mRNA sequence.  
 ACCESSION B0692817  
 VERSION B0692817.1 GI:19006075  
 KEYWORDS EST.  
 SOURCE human.  
 ORGANISM Homo sapiens  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.

REFERENCE 1 (bases 1 to 548)  
 AUTHORS Bonaldo,M.F., Lennon,G. and Soares,M.B.  
 TITLE Normalization and subtraction: two approaches to facilitate gene  
 discovery  
 JOURNAL Genome Res. 6 (9), 791-806 (1996)  
 MEDLINE 97044477  
 COMMENT  
 Contact: Soares, MB  
 Program for Rat Gene Discovery and Mapping  
 University of Iowa  
 451 Eckstein Medical Research Building Iowa City, IA 52242, USA  
 Tel: 319 335 8250  
 Fax: 319 335 9565  
 Email: msoares@blue.weeg.uiowa.edu  
 Tissue Procurement: Dr. Gregg Hageman  
 CDNA Library Preparation: Dr. M. Bento Soares, University of Iowa  
 DNA Sequencing by: Dr. M. Bento Soares, University of Iowa  
 Clone Distribution: Researchers may obtain clones from Research  
 Genetics (www.regen.com).  
 Seq primer: M13 Reverse.  
 Location/Qualifiers  
 1. 548  
 /organism="Homo sapiens"  
 /db\_xref="taxon:9606"  
 /clone\_1ib="UI-E-CKI-afk-m-06-0-UI"  
 /issue\_type="Retina Foveal and Macular"  
 /dev\_stage="adult"

## FEATURES

source  
 1. 548  
 /organism="Homo sapiens"  
 /db\_xref="taxon:9606"  
 /clone\_1ib="UI-E-CKI-afk-m-06-0-UI"  
 /issue\_type="Retina Foveal and Macular"  
 /dev\_stage="adult"



FEATURES	SOURCE
ORGANISM	human.
ORGANISM	Homo sapiens
ORGANISM	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
ORGANISM	1 (bases 1 to 529)
ORGANISM	NCI-CCAP <a href="http://www.ncbi.nlm.nih.gov/ncicgap">http://www.ncbi.nlm.nih.gov/ncicgap</a> .
ORGANISM	National Cancer Institute, Cancer Genome Anatomy Project (CGAP), Tumor Gene Index
ORGANISM	Unpublished (1997)
ORGANISM	Contact: Robert Strausberg, Ph.D.
ORGANISM	Email: <a href="mailto:cgapbs-remail.nih.gov">cgapbs-remail.nih.gov</a>
ORGANISM	Tissue Procurement: Michael J. Brownstein, M.D., Ph.D., Michael R. Emmerit-Buck, M.D., Ph.D.
ORGANISM	CDNA Library Preparation: M. Bento Soares, Ph.D.
ORGANISM	CDNA Library Arrayed by: Greg Lennon, Ph.D.
ORGANISM	DNA sequencing by: Washington University Genome Sequencing Center
ORGANISM	Clone distribution: NCI-CCAP clone distribution information can be found through the I.M.A.G.E. Consortium/LNLT at: <a href="http://www.bio.linn.gov/bdtp/image/image.html">www.bio.linn.gov/bdtp/image/image.html</a>
ORGANISM	Insert Length: 912 Std Error: 0.00
ORGANISM	Seq primer: -400P from Gibco
ORGANISM	High quality sequence stcd: 455.
ORGANISM	Location/Qualifiers
ORGANISM	1..529
ORGANISM	/organism="Homo sapiens"
ORGANISM	/db_xref="taxon:9606"
ORGANISM	/clone="IMAGE:2252679"



GenCore version 5.1.6  
Copyright (c) 1993 - 2003 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: June 29, 2003, 04:56:03 ; Search time 2476 Seconds

(without alignments)  
16420.283 Million cell updates/sec

Title: US-09-989-919-15

Perfect score: 1397

Sequence: 1 ggtcgtcactctacgcgga.....aaaaaaaaaaagcgctc 1397

Scoring table: OLIGO\_NUC

Gapop 60.0 , Gapext 60.0

Searched: 2054640 seqs, 14551402878 residues

Word size : 0

Total number of hits satisfying chosen parameters: 4109280

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Listing first 1000 summaries

Database : GenBankl:\*  
1: gb\_ba:\*  
2: gb\_hgt:\*  
3: gb\_in:\*  
4: gb\_om:\*  
5: gb\_ov:\*  
6: gb\_pac:\*  
7: gb\_ph:\*  
8: gb\_pl:\*  
9: gb\_pr:\*  
10: gb\_ro:\*  
11: gb\_scs:\*  
12: gb\_sy:\*  
13: gb\_un:\*  
14: gb\_vl:\*  
15: em\_ba:\*  
16: em\_fun:\*  
17: em\_hum:\*  
18: em\_in:\*  
19: em\_mu:\*  
20: em\_om:\*  
21: em\_or:\*  
22: em\_ov:\*  
23: em\_pat:\*  
24: em\_ph:\*  
25: em\_pl:\*  
26: em\_ro:\*  
27: em\_scs:\*  
28: em\_un:\*  
29: em\_vl:\*  
30: em\_hgt\_hum:\*  
31: em\_hgt\_inv:\*  
32: em\_hgt\_other:\*  
33: em\_hgt\_mus:\*  
34: em\_hgt\_pin:\*  
35: em\_hgt\_rcd:\*  
36: em\_hgt\_mam:\*  
37: em\_hgt\_vrt:\*  
38: em\_sy:\*  
39: em\_hgt\_hum:\*  
40: em\_hgt\_mus:\*  
41: em\_hgt\_other:\*

score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	1397	100.0	1397	6 AX480725	AX480725 Sequence
2	1258	90.1	128466	2 AC013480	AC013480 Homo sapi
3	876	62.7	2461	6 AX056460	AX056460 Sequence
4	867	62.1	1493	6 BC007901	BC007901 Homo sapi
5	442	31.6	470	6 AX480724	AX480724 Sequence
6	322	23.0	427	6 AX408491	AX408491 Sequence
7	152	10.9	462	11 G21939	G21939 human STS W
8	43	3.1	94366	2 AC112468	AC112468 Rattus no
9	43	3.1	176771	2 AC122406	AC122406 Mus muscu
10	36	2.6	1586	10 BC022157	BC022157 Mus muscu
11	36	2.6	1886	6 AX056461	AX056461 Sequence
12	33	2.4	69492	2 AC111157	AC111157 Homo sapi
13	32	2.3	164796	9 AC091069	AC091069 Homo sapi
14	32	2.3	165201	2 AC011963	AC011963 Homo sapi
15	30	2.1	1416	9 HSA420562	AJ420562 Homo sapi
16	30	2.1	2457	6 AX376502	AX376502 Sequence
17	30	2.1	3278	9 HSM800650	AL080138 Homo sapi
18	30	2.1	111723	2 AC020935	AC020935 Homo sapi
19	30	2.1	127992	2 AC105036	AC105036 Homo sapi
20	30	2.1	148263	9 AC091988	AC091988 Homo sapi
21	30	2.1	186248	2 AC095218	AC095218 Rattus no
22	29	2.1	942	9 HSA406937	AJ406937 Homo sapi
23	29	2.1	1217	8 DCU38483	U38483 D. dentatus ca
24	29	2.1	1229	10 BC004689	BC004689 Mus muscu
25	29	2.1	3329	6 AX407231	AX407231 Sequence
26	29	2.1	4301	6 AX480779	AX480779 Sequence
27	29	2.1	82360	9 AF240580	AF240580 Homo sapi
28	29	2.1	143688	2 AC127169	AC127169 Medicago
29	29	2.1	145329	2 AC055850	AC055850 Homo sapi
30	29	2.1	157871	2 AC025074	AC025074 Homo sapi
31	29	2.1	167603	2 AC068261	AC068261 Homo sapi
32	29	2.1	170355	2 AC111904	AC111904 Rattus no
33	29	2.1	173029	2 AC117646	AC117646 Mus muscu
34	29	2.1	190584	2 AC103589	AC103589 Homo sapi
35	29	2.1	198231	2 AC130141	AC130141 Rattus no
36	29	2.1	203474	9 AC069307	AC069307 Homo sapi
37	28	2.0	600	8 ATY13174	Y13174 Arabidopsis
38	28	2.0	839	9 AF043534	AF043534 Arabidops
39	28	2.0	1000	9 BC011886	BC011886 Homo sapi
40	28	2.0	1141	3 AY094810	AY094810 Drosophila
41	28	2.0	1209	9 AF087873	AF087873 Homo sapi
42	28	2.0	1318	9 BC034354	BC034354 Homo sapi
43	28	2.0	1439	9 BC000550	BC000550 Homo sapi
44	28	2.0	1560	9 AK026057	AK026057 Homo sapi
45	28	2.0	1569	9 AF131844	AF131844 Homo sapi
46	28	2.0	2405	4 BTACTRA	X57084 Bovine mRNA
47	28	2.0	2466	10 AF001157	AF001157 Mesocric
48	28	2.0	2641	3 AB073673	AB073673 Bombyx mo
49	28	2.0	2796	9 AK074352	AK074352 Homo sapi
50	28	2.0	16170	6 AX346171	AX346171 Sequence
51	28	2.0	29501	9 AC010248	AC010248 Homo sapi
52	28	2.0	37661	6 AX147034	AX147034 Sequence
53	28	2.0	46535	2 AC119222	AC119222 Mus muscu
54	28	2.0	51739	4 AC095023	AC095023 Sus scrofa
55	28	2.0	55470	2 AC116989	AC116989 Dictyoste
56	28	2.0	70121	2 AC123693	AC123693 Mus muscu
57	28	2.0	70135	2 AC095482	AC095482 Rattus no
58	28	2.0	70394	2 AC023011	AC023011 Homo sapi
59	28	2.0	71192	2 AC095584	AC095584 Rattus no
60	28	2.0	74188	2 AC021315	AC021315 Homo sapi
61	28	2.0	75002	2 AC023384	AC023384 Homo sapi
62	28	2.0	81276	2 HS276A23	AL449214 Homo sapi
63	28	2.0	92705	9 AC006327	AC006327 Homo sapi
64	28	2.0	98778	2 AC002408	AC002408 Homo sapi
65	28	2.0	102116	2 AC122681	AC122681 Rattus no

Pred. No. is the number of results predicted by chance to have a

66	C	2.0	104030	9	AC018628	AC018628 Homo sapi	C	139	2.0	199927	9	AC007055	AC007055 Homo sapi
67	C	2.0	104147	9	AC004703	AC004703 Homo sapi	C	140	2.0	200704	9	AC009781	AC009781 Homo sapi
68	C	2.0	110000	2	AL583823_0	AL583823 Homo sapi	C	141	2.0	200969	2	AC093355	AC093355 Mus muscu
69	C	2.0	110000	2	AC069419-2	Continuation (3 of	C	142	2.0	201981	2	AC130462	AC130462 Homo sapi
70	C	2.0	115762	2	AP000663	AP000663 Homo sapi	C	143	2.0	206143	2	AC106408	AC106408 Homo sapi
71	C	2.0	117864	2	AC096950	AC096950 Homo sapi	C	144	2.0	206631	2	AC005798	AC005798 Homo sapi
72	C	2.0	120269	2	AL356737	AL356737 Human DNA	C	145	2.0	210376	2	AL136118	AL136118 Homo sapi
73	C	2.0	121694	2	AC128572	AC128572 Rattus no	C	146	2.0	212971	2	AC094409	AC094409 Rattus no
74	C	2.0	125243	2	AC130997	AC130997 Mus muscu	C	147	2.0	213555	2	AC113540	AC113540 Mus muscu
75	C	2.0	134025	2	AC107697	AC107697 Rattus no	C	148	2.0	220914	2	AL731767	AL731767 Mus muscu
76	C	2.0	134778	2	AC026688	AC026688 Homo sapi	C	149	2.0	240914	2	AC127578	AC127578 Homo sapi
77	C	2.0	137068	2	AC118314	AC118314 Rattus no	C	150	2.0	244382	2	AC125044	AC125044 Homo sapi
78	C	2.0	143800	2	AC011875	AC011875 Homo sapi	C	151	2.0	260341	3	CE111648C	CE111648C Homo sapi
79	C	2.0	149195	2	AC107134	AC107134 Rattus no	C	152	2.0	264746	2	AL772379	AL772379 Mus muscu
80	C	2.0	153872	2	AC125735	AC125735 Rattus no	C	153	2.0	270438	2	AL772379	AL772379 Mus muscu
81	C	2.0	156602	2	AC012487	AC012487 Homo sapi	C	154	1.9	418	4	BFCOX7A	BFCOX7A Homo sapi
82	C	2.0	157391	9	AC007606	AC007606 Homo sapi	C	155	1.9	471	6	AX381201	AX381201 Homo sapi
83	C	2.0	160356	9	AC009068	AC009068 Homo sapi	C	156	1.9	702	2	BC001634	BC001634 Homo sapi
84	C	2.0	160785	10	AC009287	AC009287 Mus muscu	C	157	1.9	704	6	AX015526	AX015526 Homo sapi
85	C	2.0	162338	2	AC067895	AL611950 Mouse DNA	C	158	1.9	740	6	AX015526	AX015526 Homo sapi
86	C	2.0	162973	2	AC016307	AC016307 Homo sapi	C	159	1.9	757	9	HSU82613	HSU82613 Homo sapi
87	C	2.0	163621	9	AC112129	AC112129 Homo sapi	C	160	1.9	757	9	HSU82613	HSU82613 Homo sapi
88	C	2.0	163621	9	AC112129	AC112129 Homo sapi	C	161	1.9	938	6	AX048778	AX048778 Homo sapi
89	C	2.0	164820	2	AC109420	AC109420 Rattus no	C	162	1.9	1015	9	TROEST081	TROEST081 Homo sapi
90	C	2.0	165056	2	AC022412	AC022412 Homo sapi	C	163	1.9	1055	9	BC003588	BC003588 Homo sapi
91	C	2.0	165337	2	AC096212	AC096212 Rattus no	C	164	1.9	1064	9	BC009905	BC009905 Homo sapi
92	C	2.0	165929	2	AC073200	AC073200 Homo sapi	C	165	1.9	1142	6	AX008890	AX008890 Homo sapi
93	C	2.0	166625	2	AC111624	AC111624 Rattus no	C	166	1.9	1142	6	AX008890	AX008890 Homo sapi
94	C	2.0	166641	2	AC126966	AC126966 Rattus no	C	167	1.9	1441	6	AX480770	AX480770 Homo sapi
95	C	2.0	167102	9	AC084017	AC084017 Homo sapi	C	168	1.9	1465	9	BC013584	BC013584 Homo sapi
96	C	2.0	169740	2	AF186193	AF186193 Homo sapi	C	169	1.9	1543	10	AF022770	AF022770 Homo sapi
97	C	2.0	171018	2	AC125718	AC125718 Rattus no	C	170	1.9	1619	9	AF203375	AF203375 Homo sapi
98	C	2.0	171529	2	AC128417	AC128417 Rattus no	C	171	1.9	1657	10	BC021884	BC021884 Homo sapi
99	C	2.0	172848	2	AP001532	AP001532 Homo sapi	C	172	1.9	1778	9	BC015741	BC015741 Homo sapi
100	C	2.0	173331	2	AC083903	AC083903 Homo sapi	C	173	1.9	1817	6	AR070212	AR070212 Homo sapi
101	C	2.0	173331	2	AC112689	AC112689 Mus muscu	C	174	1.9	1817	6	AX330731	AX330731 Homo sapi
102	C	2.0	173655	2	AC115770	AC115770 Homo sapi	C	175	1.9	1817	6	AX330731	AX330731 Homo sapi
103	C	2.0	173772	2	AC087862	AC087862 Mus muscu	C	176	1.9	1817	6	AX330941	AX330941 Homo sapi
104	C	2.0	174666	2	AC027778	AC027778 Homo sapi	C	177	1.9	1817	6	142402	142402 Homo sapi
105	C	2.0	174974	2	AC106843	AC106843 Mus muscu	C	178	1.9	1817	6	HSICAM3RN	HSICAM3RN Homo sapi
106	C	2.0	175195	2	AP001319	AP001319 Homo sapi	C	179	1.9	2004	8	TAGL3R	TAGL3R Homo sapi
107	C	2.0	175838	9	AC026427	AC026427 Homo sapi	C	180	1.9	2013	9	AK026684	AK026684 Homo sapi
108	C	2.0	176479	2	AC096404	AC096404 Rattus no	C	181	1.9	2042	9	BC001736	BC001736 Homo sapi
109	C	2.0	176757	4	AC095025	AC095025 Sus scrof	C	182	1.9	2148	8	MINCCOI	MINCCOI Homo sapi
110	C	2.0	178418	2	AC027576	AC027576 Homo sapi	C	183	1.9	2154	8	NEUMTCOA	NEUMTCOA Homo sapi
111	C	2.0	179196	2	AC110731	AC110731 Mus muscu	C	184	1.9	2157	9	BC025687	BC025687 Homo sapi
112	C	2.0	179992	2	AC095165	AC095165 Rattus no	C	185	1.9	2311	9	HSAL33421	HSAL33421 Homo sapi
113	C	2.0	180590	2	AC106758	AC106758 Homo sapi	C	186	1.9	2354	9	HSU48437	HSU48437 Homo sapi
114	C	2.0	181583	2	AC105589	AC105589 Rattus no	C	187	1.9	2552	3	ODI308494	ODI308494 Homo sapi
115	C	2.0	181583	2	CNS07YOR	AL691403 Human chr	C	188	1.9	2585	9	BC017445	BC017445 Homo sapi
116	C	2.0	181848	2	AC112631	AC112631 Rattus no	C	189	1.9	2592	9	BC006318	BC006318 Homo sapi
117	C	2.0	183544	2	AC102571	AC102571 Mus muscu	C	190	1.9	2752	6	AX443231	AX443231 Homo sapi
118	C	2.0	184267	2	AC078796	AC078796 Homo sapi	C	191	1.9	2803	6	AX443231	AX443231 Homo sapi
119	C	2.0	184395	2	AC109983	AC109983 Rattus no	C	192	1.9	2882	6	HSAA487964	HSAA487964 Homo sapi
120	C	2.0	185704	9	AC123908	AC123908 Homo sapi	C	193	1.9	2898	6	AX443016	AX443016 Homo sapi
121	C	2.0	185721	2	AC019108	AC019108 Homo sapi	C	194	1.9	2902	9	HSAA243500	HSAA243500 Homo sapi
122	C	2.0	185846	2	AC027783	AC027783 Homo sapi	C	195	1.9	2902	9	HSAA243500	HSAA243500 Homo sapi
123	C	2.0	186340	2	AC016684	AC016684 Homo sapi	C	196	1.9	2908	6	AX443234	AX443234 Homo sapi
124	C	2.0	186367	2	AC026044	AC026044 Homo sapi	C	197	1.9	2913	6	AX443234	AX443234 Homo sapi
125	C	2.0	186962	2	AP000848	AP000848 Homo sapi	C	198	1.9	3073	6	AX443232	AX443232 Homo sapi
126	C	2.0	187099	9	AC018673	AC018673 Homo sapi	C	199	1.9	3327	6	AX472961	AX472961 Homo sapi
127	C	2.0	190000	2	AC006425	AC006425 Homo sapi	C	200	1.9	3780	6	AX253084	AX253084 Homo sapi
128	C	2.0	190544	9	AC005768	AC005768 Homo sapi	C	201	1.9	4141	6	HSMB803398	HSMB803398 Homo sapi
129	C	2.0	191328	2	AC130111	AC130111 Rattus no	C	202	1.9	4174	6	AX253082	AX253082 Homo sapi
130	C	2.0	191565	9	HUAC004531	HUAC004531 Homo sapi	C	203	1.9	5520	6	AX430963	AX430963 Homo sapi
131	C	2.0	192185	2	AC129753	AC129753 Rattus no	C	204	1.9	5605	6	HSMB804697	HSMB804697 Homo sapi
132	C	2.0	192515	2	AC117207	AC117207 Homo sapi	C	205	1.9	5893	6	AX345760	AX345760 Homo sapi
133	C	2.0	193516	2	AL135915	AL135915 Homo sapi	C	206	1.9	6106	6	AX281411	AX281411 Homo sapi
134	C	2.0	196259	2	AC097629	AC097629 Sus scrof	C	207	1.9	6106	6	AX345978	AX345978 Homo sapi
135	C	2.0	196700	2	AC097629	AC097629 Sus scrof	C	208	1.9	6106	6	AX345978	AX345978 Homo sapi
136	C	2.0	198146	2	AC021387	AC021387 Homo sapi	C	209	1.9	7426	6	AX377156	AX377156 Homo sapi
137	C	2.0	199016	2	AC021387	AC021387 Homo sapi	C	210	1.9	8675	4	SSC251837	SSC251837 Homo sapi
138	C	2.0	199016	2	AC021387	AC021387 Homo sapi	C	211	1.9	9539	6	AX277889	AX277889 Homo sapi



C 212	27	1.9	9539	6	AX3235566	AX335566 Sequence	C 285	27	1.9	101270	9	HS483K16	AL034374 Human DNA
C 213	27	1.9	10950	6	HUMTNF2	M851745 Human tumor	C 286	27	1.9	102512	9	AL355862	AL355862 Human DNA
C 214	27	1.9	12000	6	AX347455	AX347455 Sequence	C 287	27	1.9	102926	9	AC004230	AC004230 Homo sapi
C 215	27	1.9	12000	6	AX349176	AX349176 Sequence	C 288	27	1.9	103347	9	AL136086	AL136086 Human DNA
C 216	27	1.9	14362	6	AX458532	AX458532 Sequence	C 289	27	1.9	104150	9	AC026305	AC026305 Homo sapi
C 217	27	1.9	14708	6	AX251532	AX251532 Sequence	C 290	27	1.9	104241	10	AY046056	AY046056 Mus muscu
C 218	27	1.9	14708	6	AX278059	AX278059 Sequence	C 291	27	1.9	105238	9	AC011458	AC011458 Homo sapi
C 219	27	1.9	14708	6	AX323836	AX323836 Sequence	C 292	27	1.9	106610	9	AC099679	AC099679 Homo sapi
C 220	27	1.9	14708	6	AX347147	AX347147 Sequence	C 293	27	1.9	107819	9	AC010465	AC010465 Homo sapi
C 221	27	1.9	14708	6	AX356500	AX356500 Sequence	C 294	27	1.9	110000	2	AL772337_1	Continuation of
C 222	27	1.9	15913	6	HSALADG	X64467 H. sapiens A	C 295	27	1.9	111051	2	AF322449	AF322449 Homo sapi
C 223	27	1.9	16432	2	HSALADG	AJ301616 Homo sapi	C 296	27	1.9	111079	9	HS419C19	AL035407 Human DNA
C 224	27	1.9	17720	2	AC117075	AC117075 Dictyoste	C 297	27	1.9	111642	2	AC098268	AC098268 Rattus no
C 225	27	1.9	34364	9	AC090632	AC090632 Homo sapi	C 298	27	1.9	113673	9	AL135839	AL135839 Human DNA
C 226	27	1.9	35061	9	AC096569	AC096569 Homo sapi	C 299	27	1.9	113811	9	HS4503F13	AL050337 Human DNA
C 227	27	1.9	36601	2	AC116981	AC116981 Dictyoste	C 300	27	1.9	114152	9	HSJ727T10	AL050332 Human DNA
C 228	27	1.9	38651	2	AC002306	AC002306 Homo sapi	C 301	27	1.9	114420	9	AL158076	AL158076 Human DNA
C 229	27	1.9	40428	2	AC113543	AC113543 Mus muscu	C 302	27	1.9	115602	9	HS1118D24	AL031276 Human DNA
C 230	27	1.9	41112	9	AL590333	AL590332 Human DNA	C 303	27	1.9	116762	9	AC108004	AC108004 Homo sapi
C 231	27	1.9	42265	9	AC010504	AC010504 Homo sapi	C 304	27	1.9	117032	9	AC008973	AC008973 Homo sapi
C 232	27	1.9	43802	3	CER07F5	Z70284 Caenorhabdi	C 305	27	1.9	117824	9	AL389886	AL389886 Human DNA
C 233	27	1.9	44913	2	AC005786	AC005786 Homo sapi	C 306	27	1.9	118899	9	HS1085F17	AL035071 Human DNA
C 234	27	1.9	48198	2	AC008801	AC008801 Homo sapi	C 307	27	1.9	119794	9	AC021373	AC021373 Homo sapi
C 235	27	1.9	51228	2	AC094254	AC094254 Rattus no	C 308	27	1.9	120197	9	AL671863	AL671863 Human DNA
C 236	27	1.9	52102	2	AC100697	AC100697 Mus muscu	C 309	27	1.9	120955	9	HDAC002310	AC002310 Human Chr
C 237	27	1.9	54395	9	AL358117	AL358117 Homo sapi	C 310	27	1.9	121542	2	AC122993	AC122993 Rattus no
C 238	27	1.9	54787	9	AL358117	AL358117 Human DNA	C 311	27	1.9	122105	2	AL355998	AL355998 Homo sapi
C 239	27	1.9	55452	2	AC130790	AC130790 Pan trogl	C 312	27	1.9	122288	9	HS215D11	AL034417 Human DNA
C 240	27	1.9	56127	2	AC100393	AC100393 Mus muscu	C 313	27	1.9	122589	9	AL359384	AL359384 Human DNA
C 241	27	1.9	58001	9	AP001363	AP001363 Homo sapi	C 314	27	1.9	124289	9	AC010285	AC010285 Homo sapi
C 242	27	1.9	58700	2	AC103732	AC103732 Homo sapi	C 315	27	1.9	125146	2	AL158138	AL158138 Homo sapi
C 243	27	1.9	60153	2	AC117622	AC117622 Mus muscu	C 316	27	1.9	125146	9	AL512770	AL512770 Human DNA
C 244	27	1.9	60931	2	AC105113	AC105113 Homo sapi	C 317	27	1.9	126380	2	AP000714	AP000714 Homo sapi
C 245	27	1.9	62009	2	AC101298	AC101298 Mus muscu	C 318	27	1.9	126976	9	AC037473	AC037473 Homo sapi
C 246	27	1.9	62022	2	AC090301	AC090301 Homo sapi	C 319	27	1.9	127315	2	AC105109	AC105109 Homo sapi
C 247	27	1.9	62496	9	AL158836	AL158836 Human DNA	C 320	27	1.9	128736	9	AC011453	AC011453 Homo sapi
C 248	27	1.9	62496	9	AL450339	AL450339 Human DNA	C 321	27	1.9	129293	9	AC092289	AC092289 Homo sapi
C 249	27	1.9	65138	9	AL590812	AL590812 Human DNA	C 322	27	1.9	129559	9	AL731563	AL731563 Human DNA
C 250	27	1.9	66854	2	AC131266	AC131266 Homo sapi	C 323	27	1.9	131166	2	AL174866	AL174866 Homo sapi
C 251	27	1.9	67609	2	AC074083	AC074083 Homo sapi	C 324	27	1.9	131310	2	AC114893	AC114893 Felis cat
C 252	27	1.9	68206	2	AC087450	AC087450 Homo sapi	C 325	27	1.9	131716	2	AC016246	AC016246 Homo sapi
C 253	27	1.9	68297	2	AC119871	AC119871 Mus muscu	C 326	27	1.9	132668	2	AC099829	AC099829 Homo sapi
C 254	27	1.9	68662	9	HSBJ468N4	AL121752 Human DNA	C 327	27	1.9	133687	2	AC002093	AC002093 Homo sapi
C 255	27	1.9	69534	2	AC103514	AC103514 Rattus no	C 328	27	1.9	133687	9	AC011488	AC011488 Homo sapi
C 256	27	1.9	70485	9	AC104411	AC104411 Homo sapi	C 329	27	1.9	134725	2	AC091852	AC091852 Homo sapi
C 257	27	1.9	71120	2	AC124800	AC124800 Homo sapi	C 330	27	1.9	134970	9	AC002504	AC002504 Human Xp2
C 258	27	1.9	71627	9	AC004851	AC004851 Homo sapi	C 331	27	1.9	135203	9	AC025278	AC025278 Homo sapi
C 259	27	1.9	73031	2	AC101159	AC101159 Mus muscu	C 332	27	1.9	136188	2	HSBJ53A19	AL066819 Human DNA
C 260	27	1.9	73031	9	HS78B3	Z82217 Human DNA s	C 333	27	1.9	136292	2	AC116850	AC116850 Mus muscu
C 261	27	1.9	74085	9	AC107800	AC107800 Homo sapi	C 334	27	1.9	136431	9	AC093524	AC093524 Homo sapi
C 262	27	1.9	82061	9	AC025180	AC025180 Homo sapi	C 335	27	1.9	137718	9	AL137066	AL137066 Human DNA
C 263	27	1.9	82115	9	AC092159	AC092159 Homo sapi	C 336	27	1.9	138295	2	AC024448	AC024448 Homo sapi
C 264	27	1.9	83528	9	AC073254	AC073254 Homo sapi	C 337	27	1.9	139029	2	AC095197	AC095197 Rattus no
C 265	27	1.9	84231	9	AL157696	AL157696 Human DNA	C 338	27	1.9	139416	2	AC068963	AC068963 Homo sapi
C 266	27	1.9	84234	9	AC107975	AC107975 Homo sapi	C 339	27	1.9	139773	2	AC067806	AC067806 Homo sapi
C 267	27	1.9	84679	2	AC022571	AC022571 Homo sapi	C 340	27	1.9	140073	9	CNS01DPO	AL12819 Human chr
C 268	27	1.9	86298	9	AL355347	AL355347 Human DNA	C 341	27	1.9	140226	9	AC010092	AC010092 Homo sapi
C 269	27	1.9	86720	9	AL139424	AL139424 Human DNA	C 342	27	1.9	140425	9	AC005479	AC005479 Homo sapi
C 270	27	1.9	89453	9	AC068546	AC068546 Homo sapi	C 343	27	1.9	140734	9	AC025457	AC025457 Homo sapi
C 271	27	1.9	89690	9	AC106864	AC106864 Homo sapi	C 344	27	1.9	141029	2	AC122704	AC122704 Homo sapi
C 272	27	1.9	89840	2	AC122075	AC122075 Rattus no	C 345	27	1.9	141272	2	AP001976	AP001976 Homo sapi
C 273	27	1.9	90292	2	AC097539	AC097539 Rattus no	C 346	27	1.9	141388	2	AC116264	AC116264 Rattus no
C 274	27	1.9	91151	2	AC107349	AC107349 Rattus no	C 347	27	1.9	141597	2	AC008455	AC008455 Homo sapi
C 275	27	1.9	91503	2	AC021939	AC021939 Homo sapi	C 348	27	1.9	141794	2	AC093455	AC093455 Homo sapi
C 276	27	1.9	93339	9	AL161937	AL161937 Human DNA	C 349	27	1.9	141855	2	AC103905	AC103905 Homo sapi
C 277	27	1.9	95565	9	AC000379	AC000379 Human Chr	C 350	27	1.9	142037	2	AC025171	AC025171 Homo sapi
C 278	27	1.9	98250	2	AF188032	AF188032 Homo sapi	C 351	27	1.9	142298	10	AL731794	AL731794 Mouse DNA
C 279	27	1.9	98985	2	AL139427	AL139427 Homo sapi	C 352	27	1.9	142657	2	AC008773	AC008773 Homo sapi
C 280	27	1.9	99500	6	AR211954	AR211954 Sequence	C 353	27	1.9	143176	2	AC011332	AC011332 Homo sapi
C 281	27	1.9	99782	2	AC106548	AC106548 Rattus no	C 354	27	1.9	143618	2	AL137856	AL137856 Human DNA
C 282	27	1.9	99799	9	AL445184	AL445184 Human DNA	C 355	27	1.9	143618	2	AC124248	AC124248 Homo sapi
C 283	27	1.9	100259	9	AC005800	AC005800 Homo sapi	C 356	27	1.9	143812	2	AC021847	AC021847 Homo sapi
C 284	27	1.9	100902	2	AC097545	AC097545 Rattus no	C 357	27	1.9	145189	2	AC021968	AC021968 Homo sapi



C 358	27	1.9 145206	9	AC026170	431	27	1.9 164055	2	AC109719	AC109719	Rattus no
C 359	27	1.9 145459	9	AC010748	432	27	1.9 164132	2	AC019284	AC019284	Homo sapi
C 360	27	1.9 146162	2	AC112991	433	27	1.9 164679	2	AC005921	AC005921	Homo sapi
C 361	27	1.9 146860	2	AC011704	434	27	1.9 164950	2	AC022262	AC022262	Homo sapi
C 362	27	1.9 147128	2	AC127512	435	27	1.9 165028	2	AC110758	AC110758	Homo sapi
C 363	27	1.9 147131	9	AC010974	436	27	1.9 165072	2	AC023993	AC023993	Homo sapi
C 364	27	1.9 147150	2	AC120803	437	27	1.9 165490	2	AC025326	AC025326	Homo sapi
C 365	27	1.9 147885	2	AL353793	438	27	1.9 165496	2	AC104461	AC104461	Homo sapi
C 366	27	1.9 148507	10	AL713979	439	27	1.9 165690	2	AC058794	AC058794	Homo sapi
C 367	27	1.9 148911	2	AL353142	440	27	1.9 165727	2	AC018732	AC018732	Homo sapi
C 368	27	1.9 149824	2	AC009919	441	27	1.9 165947	2	AC015897	AC015897	Homo sapi
C 369	27	1.9 150193	2	AC027360	442	27	1.9 166335	2	AC010313	AC010313	Homo sapi
C 370	27	1.9 150652	2	AC130793	443	27	1.9 166418	2	AC009034	AC009034	Homo sapi
C 371	27	1.9 150804	9	AC018500	444	27	1.9 166418	9	AC009034	AC009034	Homo sapi
C 372	27	1.9 150834	9	AC005399	445	27	1.9 166439	2	AL591722	AL591722	Homo sapi
C 373	27	1.9 152186	2	AC084284	446	27	1.9 166464	9	AL158212	AL158212	Homo sapi
C 374	27	1.9 152375	2	AC123308	447	27	1.9 166847	2	AC008393	AC008393	Homo sapi
C 375	27	1.9 153094	2	AC096715	448	27	1.9 166849	2	AC127842	AC127842	Rattus no
C 376	27	1.9 153300	2	AC011933	449	27	1.9 167049	2	AC009444	AC009444	Homo sapi
C 377	27	1.9 153442	2	AC091101	450	27	1.9 167241	9	AC093114	AC093114	Homo sapi
C 378	27	1.9 153693	2	AC010373	451	27	1.9 167496	9	AL391989	AL391989	Homo sapi
C 379	27	1.9 153909	2	AC053521	452	27	1.9 167698	2	AC107050	AC107050	Homo sapi
C 380	27	1.9 154379	2	AC105661	453	27	1.9 168085	9	AC027687	AC027687	Homo sapi
C 381	27	1.9 154471	2	AC022523	454	27	1.9 168114	9	AL354828	AL354828	Homo sapi
C 382	27	1.9 154541	2	AC120212	455	27	1.9 168285	9	AC110015	AC110015	Homo sapi
C 383	27	1.9 154708	2	AC024443	456	27	1.9 168389	9	AC110015	AC110015	Homo sapi
C 384	27	1.9 154937	2	AP002382	457	27	1.9 168426	2	AC101611	AC101611	Homo sapi
C 385	27	1.9 155084	9	AC008753	458	27	1.9 168937	2	AC103233	AC103233	Homo sapi
C 386	27	1.9 155085	2	AC025620	459	27	1.9 169103	2	AC078952	AC078952	Homo sapi
C 387	27	1.9 155127	2	AC008532	460	27	1.9 169296	9	AC022198	AC022198	Homo sapi
C 388	27	1.9 155357	2	AC118923	461	27	1.9 169319	2	AC012106	AC012106	Homo sapi
C 389	27	1.9 155375	2	AC098854	462	27	1.9 169624	2	AC024187	AC024187	Homo sapi
C 390	27	1.9 155459	2	AC016619	463	27	1.9 170064	2	AC093507	AC093507	Homo sapi
C 391	27	1.9 155567	9	AC073492	464	27	1.9 170071	2	AC084733	AC084733	Homo sapi
C 392	27	1.9 155645	2	AC021153	465	27	1.9 170240	9	AL160279	AL160279	Homo sapi
C 393	27	1.9 155745	2	AC098267	466	27	1.9 170591	2	AC022936	AC022936	Homo sapi
C 394	27	1.9 155770	2	AL158162	467	27	1.9 170956	2	AC113561	AC113561	Homo sapi
C 395	27	1.9 156525	2	AC040966	468	27	1.9 170810	2	AC097851	AC097851	Homo sapi
C 396	27	1.9 156561	2	AC023113	469	27	1.9 170936	2	AL592432	AL592432	Homo sapi
C 397	27	1.9 157131	2	AC033991	470	27	1.9 171031	2	AC104020	AC104020	Homo sapi
C 398	27	1.9 157248	9	AC002312	471	27	1.9 171441	2	AC102250	AC102250	Homo sapi
C 399	27	1.9 157321	9	AC068811	472	27	1.9 171455	9	AP005208	AP005208	Homo sapi
C 400	27	1.9 157814	2	AC114310	473	27	1.9 171480	9	AC008837	AC008837	Homo sapi
C 401	27	1.9 157861	2	AC037460	474	27	1.9 171490	2	AC068209	AC068209	Homo sapi
C 402	27	1.9 158274	2	AC022767	475	27	1.9 171699	2	AC120031	AC120031	Homo sapi
C 403	27	1.9 158482	2	AC025348	476	27	1.9 171733	2	AC022626	AC022626	Homo sapi
C 404	27	1.9 158606	9	AL353805	477	27	1.9 171936	2	AC008518	AC008518	Homo sapi
C 405	27	1.9 158787	9	AC010886	478	27	1.9 172200	2	AC022626	AC022626	Homo sapi
C 406	27	1.9 159569	9	AP001318	479	27	1.9 172551	9	AC114587	AC114587	Mus muscu
C 407	27	1.9 159582	2	AC015734	480	27	1.9 172840	2	AC009127	AC009127	Homo sapi
C 408	27	1.9 159601	2	AC068167	481	27	1.9 173222	2	AC021918	AC021918	Homo sapi
C 409	27	1.9 159840	2	AP000869	482	27	1.9 173236	2	CNS07ECU	AL442183	Human chr
C 410	27	1.9 159998	2	AC074292	483	27	1.9 173260	9	AC025853	AC025853	Homo sapi
C 411	27	1.9 160628	2	AC067756	484	27	1.9 173679	2	AC007616	AC007616	Homo sapi
C 412	27	1.9 160682	2	AC026250	485	27	1.9 173852	2	AC095235	AC095235	Homo sapi
C 413	27	1.9 160766	9	AC022679	486	27	1.9 174133	2	AC104010	AC104010	Homo sapi
C 414	27	1.9 161013	9	AC018709	487	27	1.9 174361	9	AL158044	AL158044	Human DNA
C 415	27	1.9 161261	9	CNS05TE3	488	27	1.9 174831	9	AC008966	AC008966	Homo sapi
C 416	27	1.9 161658	9	AC008505	489	27	1.9 175303	9	AC009087	AC009087	Homo sapi
C 417	27	1.9 161685	2	AC103901	490	27	1.9 175332	9	AL355677	AL355677	Human sapi
C 418	27	1.9 161753	2	AF186997	491	27	1.9 175346	2	AC102746	AC102746	Mus muscu
C 419	27	1.9 162075	9	AC092581	492	27	1.9 175369	2	AC025904	AC025904	Homo sapi
C 420	27	1.9 162191	9	AC109458	493	27	1.9 175424	2	AC105435	AC105435	Sus scrofa
C 421	27	1.9 162258	9	AC105748	494	27	1.9 175499	2	AC115255	AC115255	Rattus no
C 422	27	1.9 162337	2	AC023999	495	27	1.9 175666	9	AL356481	AL356481	Human DNA
C 423	27	1.9 162477	2	AC103957	496	27	1.9 175694	2	AC016753	AC016753	Human sapi
C 424	27	1.9 162780	9	AC090445	497	27	1.9 175750	2	AC026297	AC026297	Homo sapi
C 425	27	1.9 162784	9	AC117529	498	27	1.9 175939	2	AL391872	AL391872	Human DNA
C 426	27	1.9 162975	2	AC025225	500	27	1.9 175952	9	AP002407	AP002407	Homo sapi
C 427	27	1.9 163016	2	AC115210	501	27	1.9 176292	9	AL611922	AL611922	Human DNA
C 428	27	1.9 163035	2	AC006405	502	27	1.9 176612	9			
C 429	27	1.9 163168	2	AC107047	503	27					
C 430	27	1.9 163420	2	AC012576		27					

504	27	1.9	176630	2	AC010684	AC010684 Homo sapi	577	27	1.9	192658	2	AC116702	AC116702 Mus muscu
C 505	27	1.9	176849	2	AC092145	AC092145 Homo sapi	C 578	27	1.9	193004	2	AC112390	AC112390 Rattus no
C 506	27	1.9	176861	2	AC097065	AC097065 Homo sapi	C 579	27	1.9	193157	2	AC073853	AC073853 Homo sapi
C 507	27	1.9	177028	2	AC025043	AC025043 Homo sapi	C 580	27	1.9	193424	2	AC005690	AC005690 Homo sapi
C 508	27	1.9	177449	9	AC106800	AC106800 Homo sapi	C 581	27	1.9	193509	2	AC026868	AC026868 Homo sapi
C 509	27	1.9	177638	2	AC091756	AC091756 Sus scrof	C 582	27	1.9	193519	2	AC023123	AC023123 Homo sapi
C 510	27	1.9	177824	2	AC113278	AC113278 Mus muscu	C 583	27	1.9	194339	2	AC126795	AC126795 Homo sapi
C 511	27	1.9	177848	2	AC011668	AC011668 Homo sapi	C 584	27	1.9	194462	2	AC101700	AC101700 Mus muscu
C 512	27	1.9	177962	9	AC008763	AC008763 Homo sapi	C 585	27	1.9	195297	2	AC126206	AC126206 Rattus no
C 513	27	1.9	178168	2	AC024412	AC024412 Homo sapi	C 586	27	1.9	195371	2	AC127881	AC127881 Rattus no
C 514	27	1.9	178508	2	AC023040	AC023040 Homo sapi	C 587	27	1.9	195929	10	AC098879	AC098879 Mus muscu
C 515	27	1.9	178779	2	AC117442	AC117442 Homo sapi	C 588	27	1.9	196170	2	AC087527	AC087527 Homo sapi
C 516	27	1.9	178796	2	AC020667	AC020667 Homo sapi	C 589	27	1.9	197215	9	AP002954	AP002954 Homo sapi
C 517	27	1.9	179056	2	AC068936	AC068936 Homo sapi	C 590	27	1.9	197676	2	AC022544	AC022544 Homo sapi
C 518	27	1.9	179080	2	AC092514	AC092514 Papio cyn	C 591	27	1.9	197811	2	AC009833	AC009833 Homo sapi
C 519	27	1.9	179186	2	AL357113	AL357113 Homo sapi	C 592	27	1.9	198410	2	AP000831	AP000831 Homo sapi
C 520	27	1.9	179245	9	HUAC002492	AC002492 Homo sapi	C 593	27	1.9	198418	9	AC007486	AC007486 Homo sapi
C 521	27	1.9	179455	2	AC102304	AC102304 Mus muscu	C 594	27	1.9	198542	2	AC023087	AC023087 Homo sapi
C 522	27	1.9	180096	2	AC025483	AC025483 Homo sapi	C 595	27	1.9	198891	2	AL845291	AL845291 Mus muscu
C 523	27	1.9	180306	2	AC024701	AC024701 Homo sapi	C 596	27	1.9	199284	2	AC107388	AC107388 Homo sapi
C 524	27	1.9	180377	2	AC126229	AC126229 Papio cyn	C 597	27	1.9	200059	2	AC108837	AC108837 Mus muscu
C 525	27	1.9	180538	2	AC073514	AC073514 Homo sapi	C 598	27	1.9	200147	9	AC023424	AC023424 Homo sapi
C 526	27	1.9	181159	2	AC112830	AC112830 Rattus no	C 599	27	1.9	200336	2	AC112955	AC112955 Mus muscu
C 527	27	1.9	181413	2	AP001352	AP001352 Homo sapi	C 600	27	1.9	200618	9	AC009061	AC009061 Homo sapi
C 528	27	1.9	181650	2	AC097825	AC097825 Rattus no	C 601	27	1.9	200791	2	AC125507	AC125507 Papio cyn
C 529	27	1.9	181707	2	AC027720	AC027720 Homo sapi	C 602	27	1.9	201157	2	AC079189	AC079189 Homo sapi
C 530	27	1.9	181926	2	AC008461	AC008461 Homo sapi	C 603	27	1.9	201466	2	AC012335	AC012335 Homo sapi
C 531	27	1.9	182205	2	AC104478	AC104478 Sus scrof	C 604	27	1.9	201546	2	AL772224	AL772224 Mus muscu
C 532	27	1.9	182452	9	AL161731	AL161731 Human DNA	C 605	27	1.9	201583	2	AC122358	AC122358 Mus muscu
C 533	27	1.9	182726	2	AC122192	AC122192 Mus muscu	C 606	27	1.9	201724	2	AC119506	AC119506 Rattus no
C 534	27	1.9	182789	9	AC025446	AC025446 Homo sapi	C 607	27	1.9	201994	2	AC113122	AC113122 Mus muscu
C 535	27	1.9	182988	2	AC021017	AC021017 Homo sapi	C 608	27	1.9	202143	2	AC096041	AC096041 Rattus no
C 536	27	1.9	183176	9	AC106894	AC106894 Homo sapi	C 609	27	1.9	202235	2	AC117378	AC117378 Homo sapi
C 537	27	1.9	183604	2	AC022022	AC022022 Homo sapi	C 610	27	1.9	202501	2	AC114600	AC114600 Mus muscu
C 538	27	1.9	183652	2	AC009611	AC009611 Homo sapi	C 611	27	1.9	202727	2	AC109724	AC109724 Rattus no
C 539	27	1.9	183800	2	AC123118	AC123118 Rattus no	C 612	27	1.9	202772	2	AC109724	AC109724 Rattus no
C 540	27	1.9	184130	2	AC092141	AC092141 Homo sapi	C 613	27	1.9	202773	2	AC025220	AC025220 Homo sapi
C 541	27	1.9	184313	9	AL445584	AL445584 Human DNA	C 614	27	1.9	202995	2	AC019134	AC019134 Homo sapi
C 542	27	1.9	184989	2	AC113392	AC113392 Homo sapi	C 615	27	1.9	203282	2	AL772151	AL772151 Mus muscu
C 543	27	1.9	185053	2	AC097334	AC097334 Pan trogl	C 616	27	1.9	203336	10	AL591433	AL591433 Mouse DNA
C 544	27	1.9	185219	2	AC103410	AC103410 Mus muscu	C 617	27	1.9	203537	2	AC008734	AC008734 Homo sapi
C 545	27	1.9	185321	2	AC026153	AC026153 Homo sapi	C 618	27	1.9	203572	2	AC016769	AC016769 Homo sapi
C 546	27	1.9	185373	2	CNS0180X	AL109767 Homo sapi	C 619	27	1.9	203579	2	AL645762	AL645762 Homo sapi
C 547	27	1.9	185399	2	AC026940	AC026940 Homo sapi	C 620	27	1.9	203982	9	AC067854	AC067854 Homo sapi
C 548	27	1.9	185409	10	AC068905	AC068905 Mus muscu	C 621	27	1.9	204534	2	AC084400	AC084400 Mus muscu
C 549	27	1.9	186170	2	AC116468	AC116468 Mus muscu	C 622	27	1.9	205286	2	AC113434	AC113434 Mus muscu
C 550	27	1.9	186236	9	AC123962	AC123962 Pan trogl	C 623	27	1.9	205594	2	AP002802	AP002802 Homo sapi
C 551	27	1.9	186562	2	AL451079	AL451079 Homo sapi	C 624	27	1.9	205673	2	AC073131	AC073131 Homo sapi
C 552	27	1.9	186730	2	AC097589	AC097589 Sus scrof	C 625	27	1.9	205915	2	AC099489	AC099489 Homo sapi
C 553	27	1.9	186981	9	AC006459	AC006459 Homo sapi	C 626	27	1.9	205919	2	AC012279	AC012279 Homo sapi
C 554	27	1.9	187006	2	AC129443	AC129443 Rattus no	C 627	27	1.9	206019	2	AC078875	AC078875 Homo sapi
C 555	27	1.9	187075	2	AL445424	AL445424 Homo sapi	C 628	27	1.9	206475	9	AC010548	AC010548 Homo sapi
C 556	27	1.9	187110	2	AC084224	AC084224 Homo sapi	C 629	27	1.9	206712	9	AC104302	AC104302 Homo sapi
C 557	27	1.9	188841	2	AC120213	AC120213 Canis fam	C 630	27	1.9	207418	2	AC090774	AC090774 Homo sapi
C 558	27	1.9	188846	2	AC122428	AC122428 Mus muscu	C 631	27	1.9	207575	2	AC124438	AC124438 Mus muscu
C 559	27	1.9	188926	9	AC012182	AC012182 Homo sapi	C 632	27	1.9	208009	2	AC126365	AC126365 Homo sapi
C 560	27	1.9	188989	10	AC074224	AC074224 Mus muscu	C 633	27	1.9	208335	2	AC113206	AC113206 Mus muscu
C 561	27	1.9	189312	2	AC126406	AC126406 Homo sapi	C 634	27	1.9	209185	2	AC098292	AC098292 Rattus no
C 562	27	1.9	189396	2	AL161418	AL161418 Homo sapi	C 635	27	1.9	209262	2	AC021091	AC021091 Homo sapi
C 563	27	1.9	189476	2	AC009893	AC009893 Homo sapi	C 636	27	1.9	210164	2	AC010732	AC010732 Homo sapi
C 564	27	1.9	189579	9	AL354733	AL354733 Human DNA	C 637	27	1.9	210179	2	AC126760	AC126760 Homo sapi
C 565	27	1.9	189917	9	AP000763	AP000763 Homo sapi	C 638	27	1.9	210250	2	AC009843	AC009843 Homo sapi
C 566	27	1.9	190076	9	AC008403	AC008403 Homo sapi	C 639	27	1.9	210617	2	AC034201	AC034201 Homo sapi
C 567	27	1.9	190440	9	AL357515	AL357515 Human DNA	C 640	27	1.9	211542	2	AL672182	AL672182 Mus muscu
C 568	27	1.9	190858	9	AC007537	AC007537 Homo sapi	C 641	27	1.9	211583	2	AC069025	AC069025 Homo sapi
C 569	27	1.9	190982	2	AC025285	AC025285 Homo sapi	C 642	27	1.9	211595	2	AC019334	AC019334 Homo sapi
C 570	27	1.9	191155	9	AC022211	AC022211 Homo sapi	C 643	27	1.9	211937	2	AP008429	AP008429 Mus muscu
C 571	27	1.9	191280	2	AC087389	AC087389 Homo sapi	C 644	27	1.9	211937	2	AP003064	AP003064 Homo sapi
C 572	27	1.9	191481	2	AC108123	AC108123 Homo sapi	C 645	27	1.9	212080	2	AC091457	AC091457 Mus muscu
C 573	27	1.9	191546	2	AL355982	AL355982 Homo sapi	C 646	27	1.9	212480	2	AL606927	AL606927 Mus muscu
C 574	27	1.9	191673	2	AL356743	AL356743 Homo sapi	C 647	27	1.9	212647	10	AL606987	AL606987 Mouse DNA
C 575	27	1.9	192391	9	AC010768	AC010768 Homo sapi	C 648	27	1.9	213034	2	AC068608	AC068608 Mus muscu
C 576	27	1.9	192437	2	AC107199	AC107199 Rattus no	C 649	27	1.9	213327	2	AC122805	AC122805 Mus muscu

650	27	1.9	213361	2	AC009121	Homo sapi	723	26	1.9	528	9	AF521046	AF521046 Homo sapi
651	27	1.9	213464	2	AC022165	Homo sapi	724	26	1.9	549	6	AX019168	AX019168 Sequence
652	27	1.9	214665	2	AC096374	Rattus no	725	26	1.9	549	6	AX019716	AX019716 Sequence
653	27	1.9	214852	9	CNS05TC2	Human chr	726	26	1.9	580	6	AX186292	AX186292 Sequence
654	27	1.9	216021	9	HUAC004787	Homo sapi	727	26	1.9	594	6	AX406849	AX406849 Sequence
655	27	1.9	216387	2	DJ534K4	Homo sapi	728	26	1.9	607	3	AF273728	AF273728 Heteroder
656	27	1.9	216681	2	AC006534	Homo sapi	729	26	1.9	631	9	AF417160	AF417160 Homo sapi
657	27	1.9	217186	2	AC118472	Homo sapi	730	26	1.9	668	9	AF479317	AF479317 Homo sapi
658	27	1.9	217328	2	AC007513	Homo sapi	731	26	1.9	690	8	ATH245631	ATH245631 Arabidops
659	27	1.9	218542	2	AL731655	Mus muscu	732	26	1.9	700	6	AX182361	AX182361 Sequence
660	27	1.9	218699	2	AC125063	Mus muscu	733	26	1.9	757	8	SCYGR206W	SCYGR206W
661	27	1.9	219037	10	AL596258	Mouse DNA	734	26	1.9	772	9	AF486812	AF486812 Homo sapi
662	27	1.9	219451	2	AC125209	Mus muscu	735	26	1.9	772	9	AF521049	AF521049 Homo sapi
663	27	1.9	219752	2	AC007215	Homo sapi	736	26	1.9	784	6	AX281933	AX281933 Sequence
664	27	1.9	221467	2	AC020866	Mus muscu	737	26	1.9	796	5	AF452589	AF452589 Homo sapi
665	27	1.9	221941	9	AC091060	Homo sapi	738	26	1.9	807	5	AY034102	AY034102 Poecilia
666	27	1.9	222387	2	AC091677	Mus muscu	739	26	1.9	837	5	AY034101	AY034101 Poecilia
667	27	1.9	223597	2	AL512402	Homo sapi	740	26	1.9	840	8	AY102540	AY102540 Arabidops
668	27	1.9	223606	2	AC013479	Homo sapi	741	26	1.9	899	6	AR157928	AR157928 Sequence
669	27	1.9	223700	2	AC127365	Mus muscu	742	26	1.9	899	8	AF065885	AF065885 Nicotiana
670	27	1.9	226467	2	AC121855	Mus muscu	743	26	1.9	927	5	AX430926	AX430926 Sequence
671	27	1.9	227564	2	AC099094	Rattus no	744	26	1.9	944	8	HSU85611	HSU85611 Human DNA-P
672	27	1.9	229506	2	AL161933	Homo sapi	745	26	1.9	949	8	AF488565	AF488565 Arabidops
673	27	1.9	231039	10	AC090480	Mus Muscu	746	26	1.9	975	10	BC021849	BC021849 Mus muscu
674	27	1.9	231566	2	AL158811	Homo sapi	747	26	1.9	1040	10	AF166097	AF166097 Mus muscu
675	27	1.9	233204	10	AL596068	Mouse DNA	748	26	1.9	1068	8	AF043530	AF043530 Arabidops
676	27	1.9	233548	2	AC068757	Homo sapi	749	26	1.9	1075	5	AF133178	AF133178 Arabidops
677	27	1.9	237059	2	AL807247	Mus muscu	750	26	1.9	1075	5	AF390589	AF390589 Anas plat
678	27	1.9	241262	2	AC113973	Mus muscu	751	26	1.9	1089	10	BC019510	BC019510 Mus muscu
679	27	1.9	241278	9	AC097268	Pan trogl	752	26	1.9	1109	10	BC019710	BC019710 Mus muscu
680	27	1.9	244023	2	AL606973	Mus muscu	753	26	1.9	1126	5	AF143875	AF143875 Homo sapi
681	27	1.9	249021	9	AC105933	Homo sapi	754	26	1.9	1134	5	AB072006	AB072006 Xenopus 1
682	27	1.9	249846	2	AC125227	Mus muscu	755	26	1.9	1135	9	AF119337	AF119337 Macropus
683	27	1.9	252730	2	AC027740	Mus muscu	756	26	1.9	1188	9	AY065987	AY065987 Homo sapi
684	27	1.9	254336	10	AC098838	Genomic s	757	26	1.9	1233	9	HS4420500	HS4420500 Homo sapi
685	27	1.9	267375	2	AC115418	Rattus no	758	26	1.9	1271	10	BC030062	BC030062 Mus muscu
686	27	1.9	286332	2	AC022105	Homo sapi	759	26	1.9	1274	6	AX411627	AX411627 Sequence
687	27	1.9	287205	2	AC123865	Mus muscu	760	26	1.9	1345	5	MUSCRKNB3	M74148 Mouse creat
688	27	1.9	293545	2	AC098276	Rattus no	761	26	1.9	1406	5	CC2P311	Z48973 C. carpio fo
689	27	1.9	296820	10	AF312033	Mus muscu	762	26	1.9	1432	6	AX320875	AX320875 Sequence
690	27	1.9	309430	2	AC122209	Homo sapi	763	26	1.9	1447	9	BC000522	BC000522 Homo sapi
691	27	1.9	309945	2	AC122312	Mus muscu	764	26	1.9	1498	10	BC015271	BC015271 Mus muscu
692	27	1.9	320781	2	AC110095	Rattus no	765	26	1.9	1505	9	BC019252	BC019252 Homo sapi
693	27	1.9	331211	9	AB016897	Homo sapi	766	26	1.9	1524	3	GLU92646	U92646 Gecarcinus
694	27	1.9	331657	2	AC124735	Mus muscu	767	26	1.9	1560	9	HSM801155	AL117619 Homo sapi
695	27	1.9	331657	2	AC125157	Mus muscu	768	26	1.9	1568	10	BC003796	BC003796 Mus muscu
696	27	1.9	335667	2	AC094907	Rattus no	769	26	1.9	1586	10	BC034880	BC034880 Mus muscu
697	27	1.9	343300	2	AC107296	Homo sapi	770	26	1.9	1599	9	BC031557	BC031557 Homo sapi
698	27	1.9	349980	6	AX344559	Mus muscu	771	26	1.9	1621	9	BC031557	BC031557 Homo sapi
699	27	1.9	349980	6	AX344570	Mus muscu	772	26	1.9	1630	10	BC024618	BC024618 Mus muscu
700	27	1.9	349980	6	AX344570	Mus muscu	773	26	1.9	1640	10	BC003898	BC003898 Mus muscu
701	27	1.9	349980	6	AX344570	Mus muscu	774	26	1.9	1644	9	AK024737	AK024737 Homo sapi
702	27	1.9	349980	6	AX344570	Mus muscu	775	26	1.9	1657	8	SCYGR205W	Z72990 S. cerevisia
703	27	1.9	349980	6	AX344570	Mus muscu	776	26	1.9	1676	10	BC027808	BC027808 Mus muscu
704	27	1.9	349980	6	AX344570	Mus muscu	777	26	1.9	1680	9	AB060877	AB060877 Macaca fa
705	27	1.9	349980	6	AX344570	Mus muscu	778	26	1.9	1700	10	BC027427	BC027427 Mus muscu
706	27	1.9	349980	6	AX344570	Mus muscu	779	26	1.9	1748	9	BC027863	BC027863 Homo sapi
707	27	1.9	349980	6	AX344570	Mus muscu	780	26	1.9	1758	6	AS1201	AS1201 Sequence 3
708	27	1.9	349980	6	AX344570	Mus muscu	781	26	1.9	1758	6	AR112865	AR112865 Sequence
709	27	1.9	349980	6	AX344570	Mus muscu	782	26	1.9	1797	6	E41372	E41372 Amino acid
710	27	1.9	349980	6	AX344570	Mus muscu	783	26	1.9	1797	6	HSM801476	AL133614 Homo sapi
711	27	1.9	349980	6	AX344570	Mus muscu	784	26	1.9	1852	9	RNU59324	U59324 Rattus norv
712	27	1.9	349980	6	AX344570	Mus muscu	785	26	1.9	1852	9	AK024413	AK024413 Homo sapi
713	27	1.9	349980	6	AX344570	Mus muscu	786	26	1.9	1859	9	AB060211	AB060211 Macaca fa
714	27	1.9	349980	6	AX344570	Mus muscu	787	26	1.9	1862	5	AF091553	AF091553 Danio rer
715	27	1.9	349980	6	AX344570	Mus muscu	788	26	1.9	1893	9	BC012350	BC012350 Homo sapi
716	27	1.9	349980	6	AX344570	Mus muscu	789	26	1.9	1905	4	S77394	S77394 insuln-lik
717	27	1.9	349980	6	AX344570	Mus muscu	790	26	1.9	1918	8	BC015448	BC015448 Homo sapi
718	27	1.9	349980	6	AX344570	Mus muscu	791	26	1.9	1932	10	BC027172	BC027172 Mus muscu
719	27	1.9	349980	6	AX344570	Mus muscu	792	26	1.9	1958	3	AMEP1A1	X52884 A. mellif
720	27	1.9	349980	6	AX344570	Mus muscu	793	26	1.9	1970	8	ATPME1	X81585 A. thaliana
721	27	1.9	349980	6	AX344570	Mus muscu	794	26	1.9	1971	8	AK026751	AK026751 Homo sapi
722	27	1.9	349980	6	AX344570	Mus muscu	795	26	1.9	2069	9	BC018289	BC018289 Homo sapi

796	26	1.9	2083	9	BC031687	BC031687 Homo sapi	869	26	1.9	22194	9	AL589644	AL589644 Human DNA
797	26	1.9	2085	9	BC002354	BC002354 Homo sapi	870	26	1.9	22445	9	AL353669	AL353669 Human DNA
798	26	1.9	2093	6	I40967	I40967 Sequence 1	871	26	1.9	23466	10	AB089806	AB089806 Mus muscu
799	26	1.9	2093	9	CAHAVCR1	X98252 C.aethiops	872	26	1.9	23741	2	AC114397	AC114397 Typhnoso
800	26	1.9	2099	9	AK025646	AK025646 Homo sapi	873	26	1.9	24276	9	AC091800	AC091800 Homo sapi
801	26	1.9	2118	9	AK098377	AK098377 Homo sapi	874	26	1.9	25662	10	AF169829	AF169829 Mus muscu
802	26	1.9	2125	9	BC004453	BC004453 Homo sapi	875	26	1.9	26106	9	AC104534	AC104534 Homo sapi
803	26	1.9	2134	10	BC013663	BC013663 Mus muscu	876	26	1.9	31343	2	AC115576	AC115576 Dictyoste
804	26	1.9	2137	9	BC002361	BC002361 Homo sapi	877	26	1.9	31565	9	AL354660	AL354660 Homo sapi
805	26	1.9	2158	9	AF258582	AF258582 Homo sapi	878	26	1.9	32797	9	AB014088	AB014088 Human DNA
806	26	1.9	2159	10	AF172088	AF172088 Mus muscu	879	26	1.9	32887	9	AC118535	AC118535 Homo sapi
807	26	1.9	2227	9	AK000380	AK000380 Homo sapi	880	26	1.9	34008	9	AL450283	AL450283 Human DNA
808	26	1.9	2288	9	AB052196	AB052196 Macaca fa	881	26	1.9	35705	9	AC010522	AC010522 Homo sapi
809	26	1.9	2339	5	DRE489219	AJ489219 Danio rer	882	26	1.9	37194	9	HSV210B9	Z70802 Human DNA s
810	26	1.9	2369	6	AX017988	AX017988 Sequence	883	26	1.9	37307	9	HS335H7	AC0041258 Human DNA
811	26	1.9	2371	9	BC025773	BC025773 Homo sapi	884	26	1.9	37616	9	AC069404	AC069404 Homo sapi
812	26	1.9	2378	9	HSMB05223	AL833916 Homo sapi	885	26	1.9	38825	9	AC090098	AC090098 Homo sapi
813	26	1.9	2416	10	AB041652	AB041652 Mus muscu	886	26	1.9	39880	9	AC010645	AC010645 Homo sapi
814	26	1.9	2423	9	BC027937	BC027937 Homo sapi	887	26	1.9	39985	10	AL731802	AL731802 Mouse DNA
815	26	1.9	2433	10	BC016266	BC016266 Mus muscu	888	26	1.9	40587	9	AC102952	AC102952 Homo sapi
816	26	1.9	2466	9	HSMB01702	AL136734 Homo sapi	889	26	1.9	41021	9	HS0781L3	HS0781L3
817	26	1.9	2466	10	AB041574	AB041574 Mus muscu	890	26	1.9	41139	9	AL596262	AL596262 Homo sapi
818	26	1.9	2467	3	AB036857	AB036857 Clona int	891	26	1.9	41280	9	AL589906	AL589906 Human DNA
819	26	1.9	2556	9	HSMB02291	AL137553 Homo sapi	892	26	1.9	41303	9	AL353602	AL353602 Human DNA
820	26	1.9	2585	9	AB046640	AB046640 Macaca fa	893	26	1.9	41876	9	AC004190	AC004190 Homo sapi
821	26	1.9	2660	6	AX320636	AX320636 Sequence	894	26	1.9	43299	9	AC005756	AC005756 Homo sapi
822	26	1.9	2673	9	BC000137	BC000137 Homo sapi	895	26	1.9	43594	9	AC093144	AC093144 Homo sapi
823	26	1.9	2688	8	AF043085	AF043085 Lycopersi	896	26	1.9	44693	2	AC087338	AC087338 Homo sapi
824	26	1.9	2802	9	BC001665	BC001665 Homo sapi	897	26	1.9	45276	9	AL512654	AL512654 Human DNA
825	26	1.9	2837	9	BC016392	BC016392 Homo sapi	898	26	1.9	45692	9	AC108153	AC108153 Homo sapi
826	26	1.9	2963	9	AK025104	AK025104 Homo sapi	899	26	1.9	46165	2	AC092564	AC092564 Hylobates
827	26	1.9	2975	9	BC029130	BC029130 Homo sapi	900	26	1.9	46581	2	AC008982	AC008982 Homo sapi
828	26	1.9	3001	6	AX347465	AX347465 Sequence	901	26	1.9	47090	9	AC092310	AC092310 Homo sapi
829	26	1.9	3001	6	AX349186	AX349186 Sequence	902	26	1.9	47090	9	AC093233	AC093233 Homo sapi
830	26	1.9	3071	9	AK026934	AK026934 Homo sapi	903	26	1.9	47434	2	AC023918	AC023918 Homo sapi
831	26	1.9	3149	9	HSEMR1	X81479 H.sapiens m	904	26	1.9	48700	9	AP000558	AP000558 Homo sapi
832	26	1.9	3149	11	G28602	G28602 human STS S	905	26	1.9	48702	2	AC080410	AC080410 Homo sapi
833	26	1.9	3253	9	BC034376	BC034376 Homo sapi	906	26	1.9	51874	2	AC129511	AC129511 Homo sapi
834	26	1.9	3318	9	HS2PAPSS3	AF160505 Homo sapi	907	26	1.9	51874	2	AC101066	AC101066 Homo sapi
835	26	1.9	3338	6	AX046378	AX046378 Sequence	908	26	1.9	53552	2	AC100796	AC100796 Homo sapi
836	26	1.9	3408	6	AX046353	AX046353 Sequence	909	26	1.9	53844	2	AC115993	AC115993 Homo sapi
837	26	1.9	3457	3	AF310886	AF310886 Dictyoste	910	26	1.9	56085	2	AC024477	AC024477 Homo sapi
838	26	1.9	3473	9	HSMB04332	AL833021 Homo sapi	911	26	1.9	57287	9	AC008283	AC008283 Homo sapi
839	26	1.9	3891	6	AR084970	AR084970 Sequence	912	26	1.9	57671	2	AC100231	AC100231 Mus muscu
840	26	1.9	3904	10	BC019148	AR081948 Mus muscu	913	26	1.9	57711	2	AC091853	AC091853 Homo sapi
841	26	1.9	3994	10	MUSMHH2MAX	L14278 Mouse MHC C	914	26	1.9	57904	2	AC091648	AC091648 Homo sapi
842	26	1.9	4028	3	AY095056	AY095056 Drosophill	915	26	1.9	58522	9	AL445263	AL445263 Human DNA
843	26	1.9	4883	8	YSCADE3	M12878 Saccharomyc	916	26	1.9	58640	2	AC111698	AC111698 Rattus no
844	26	1.9	5647	6	AX252020	AX252020 Sequence	917	26	1.9	58723	9	AL312337	AL312337 Human DNA
845	26	1.9	5647	6	AX346468	AX346468 Sequence	918	26	1.9	58841	2	AC127914	AC127914 Rattus no
846	26	1.9	5647	6	AX348787	AX348787 Sequence	919	26	1.9	59102	2	AL133324	AL133324 Human DNA
847	26	1.9	5794	9	HSPPLGS	AF041003 Homo sapi	920	26	1.9	59245	2	AC110587	AC110587 Homo sapi
848	26	1.9	6062	9	AF181099	AF181099 Homo sapi	921	26	1.9	59991	2	AC090289	AC090289 Mus muscu
849	26	1.9	6086	6	AX251421	AX251421 Sequence	922	26	1.9	60036	2	AC009610	AC009610 Homo sapi
850	26	1.9	6165	9	AX411222	AX411222 Sequence	923	26	1.9	60725	9	AL669983	AL669983 Human DNA
851	26	1.9	6165	9	HSMTCP12A	Z24459 H.sapiens M	924	26	1.9	60965	2	AC090090	AC090090 Homo sapi
852	26	1.9	6352	6	AX345806	AX345806 Sequence	925	26	1.9	61946	2	AC116422	AC116422 Homo sapi
853	26	1.9	6484	6	AX344730	AX344730 Sequence	926	26	1.9	62709	2	AC117730	AC117730 Mus muscu
854	26	1.9	6668	6	AX346599	AX346599 Sequence	927	26	1.9	63024	2	AC129476	AC129476 Homo sapi
855	26	1.9	6686	6	AX345363	AX345363 Sequence	928	26	1.9	63334	2	AC032023	AC032023 Homo sapi
856	26	1.9	7921	6	AX346872	AX346872 Sequence	929	26	1.9	63403	10	AL611923	AL611923 Mouse DNA
857	26	1.9	9139	6	AX057394	AX057394 Sequence	930	26	1.9	63605	9	AF62787	AF62787 Homo sapi
858	26	1.9	9143	6	AR075140	AR075140 Sequence	931	26	1.9	63739	9	AF271897	AF271897 Human DNA
859	26	1.9	9143	14	HGU22304	U22304 Hepatitlis G	932	26	1.9	63750	2	AC090563	AC090563 Mus muscu
860	26	1.9	9399	6	AX057392	AX057392 Sequence	933	26	1.9	64417	2	AC129943	AC129943 Mus muscu
861	26	1.9	9399	14	AF179612	AF179612 Hepatitlis	934	26	1.9	64888	2	AC123746	AC123746 Homo sapi
862	26	1.9	10891	6	AX345366	AX345366 Sequence	935	26	1.9	65174	2	AC025214	AC025214 Homo sapi
863	26	1.9	14753	6	AR179443	AR179443 Sequence	936	26	1.9	65346	2	AC026597	AC026597 Homo sapi
864	26	1.9	16170	6	AX346170	AX346170 Sequence	937	26	1.9	65452	2	AC016247	AC016247 Homo sapi
865	26	1.9	17835	10	AF179427	AF179427 Mus muscu	938	26	1.9	65643	2	AC073386	AC073386 Homo sapi
866	26	1.9	18230	8	SCPEGK620	Z49133 S.cerevisia	939	26	1.9	65651	2	AC098425	AC098425 Rattus no
867	26	1.9	21528	4	AC012608	AC012608 Homo sapi	940	26	1.9	65874	2	AC093310	AC093310 Homo sapi
868	26	1.9	21737	2	SSTMP835	X99893 S.scrofa tm	941	26	1.9	66338	2	AC116694	AC116694 Mus muscu



```

JOURNAL      Submitted (07-NOV-2001) Department of Genetics, Washington
              University, 4444 Forest Park Avenue, St. Louis, Missouri 63108, USA
COMMENT       On Nov 7, 2001 this sequence version replaced gi:13399435.
-----
              Center: Washington University Genome Sequencing Center
              Genome Center
Center code:   WUGSC
Web site:     http://genome.wustl.edu/gsc
Contact:      saplens@watson.wustl.edu
               Summary Statistics
Center project name: H_NH0564016
-----
NOTICE: This sequence may not represent the entire insert of this clone. It may be shorter because we only sequence overlapping clone sections once, or longer because we provide a small overlap between neighboring data submissions.

This sequence was finished as follows unless otherwise noted:
All regions were double stranded, sequenced with an alternate chemistry, or covered by high quality data (i.e., phred quality >= 30); an attempt was made to resolve all sequencing problems, such as compressions and repeats; all regions were covered by sequence from more than one subclone; and the assembly was confirmed by restriction digest.

MAPPING INFORMATION:
Mapping information for this clone was provided by Dr. John D. McPherson, Department of Genetics, Washington University, St. Louis MO. For additional information about the map position of this sequence, see http://genome.wustl.edu/gsc

SOURCE INFORMATION:
The RPCI-11 human BAC library was made from the blood of one male donor, as described by Osoegawa, K., Moon, P.Y., Zhao, B., Frengen, E., Teleno, M., Catanesi, J.J. and de Jong, P.J. (1998) An improved approach for construction of bacterial artificial chromosome libraries. Genomics 51:1-8. The clone may be obtained either from Research Genetics, Inc. (http://www.resgen.com) or Pieter de Jong and coworkers at the Roswell Park Cancer Institute (http://bacpac.med.buffalo.edu)
VECTOR:        pBACec3.6

NEIGHBORING SEQUENCE INFORMATION:
The clone sequenced to the left is AC068592; the clone sequenced to the right is CTD-2096A9. Actual start of this clone is at base position 1 of RP11-564016.

The sequence from base position 103537 to 103654 and from base position 103940 to 104887 is derived from PCR of BAC DNA.

The fidelity of the sequence from base position 48371 to 48772 cannot be guaranteed due to an unresolved imperfect dinucleotide repeat.

FEATURES
Source
    Location/Qualifiers
      1..128466
         /organism="Homo sapiens"
         /db_xref="taxon:9606"
         /chromosome="2"
         /map="2"
         /clone="RP11-564016"
         /clone_id="RPCI-11"
      291..452
         /rpt_family="Alu"
      1484..1617
         /rpt_family="L2"
      1669..1826
         /rpt_family="Alu"
      1864..2412
         /rpt_family="WALR"
      2413..2591
         /rpt_family="MIR"
      2689..2833
         /rpt_family="L2"
      3055..3082
         /repeat_region

```

```
/rpt_family="(T)n"
repeat_region 3085..3392
/rpt_family="Alu"
repeat_region 3530..3643
/rpt_family="CR1"
repeat_region 3644..3717
/rpt_family="Mariner"
repeat_region 3718..4183
/rpt_family="CR1"
repeat_region 6123..6293
/rpt_family="MIR"
repeat_region 6845..6881
/rpt_family="AT_rich"
repeat_region 6911..6963
/rpt_family="AT_rich"
repeat_region 7910..8008
/rpt_family="MIR"
repeat_region 8009..8299
/rpt_family="Alu"
repeat_region 8300..8326
/rpt_family="MIR"
repeat_region 8327..8599
/rpt_family="Alu"
repeat_region 8600..8668
/rpt_family="MIR"
repeat_region 9463..9553
/rpt_family="(TTCA)n"
repeat_region 9890..9923
/rpt_family="AT_rich"
repeat_region 10066..10473
/rpt_family="L2"
repeat_region 10474..10520
/rpt_family="(CA)n"
repeat_region 10521..10746
/rpt_family="L2"
repeat_region 12200..12385
/rpt_family="MIR"
repeat_region 12452..12489
/rpt_family="(CACG)n"
repeat_region 13410..13773
/rpt_family="MaLR"
repeat_region 14076..14317
/rpt_family="MIR"
repeat_region 15470..15592
/rpt_family="MIR"
repeat_region 15599..15967
/rpt_family="MaLR"
misc_feature 16227..16617
/ncore="similar to EST AA913068 (NID:G3052460) o134e06.s1"
repeat_region 16517..16693
/rpt_family="L1"
repeat_region 16717..16788
/rpt_family="L1"
repeat_region 16988..17009
/rpt_family="AT_rich"
repeat_region 18365..18509
/rpt_family="MER103"
repeat_region 18594..18644
/rpt_family="(TG)n"
repeat_region 19032..19193
/rpt_family="MER1_type"
repeat_region 19359..19668
/rpt_family="L2"
repeat_region 22454..22498
/rpt_family="GA-rich"
repeat_region 25334..25644
/rpt_family="Alu"
repeat_region 25647..25768
/rpt_family="MIR"
repeat_region 25770..25888
/rpt_family="L2"
repeat_region 25893..26193
/rpt_family="Alu"
```

```
misc_feature 25910..26358
/ncore="similar to EST AA578847 (NID:G2357031) nk66a09.s1"
repeat_region 26368..26432
/rpt_family="MIR"
repeat_region 26502..26568
/rpt_family="MIR"
repeat_region 26725..27274
/rpt_family="MaLR"
repeat_region 27275..27400
/rpt_family="MIR"
repeat_region 28344..28444
/rpt_family="MIR"
repeat_region 28445..28737
/rpt_family="Alu"
repeat_region 28738..28877
/rpt_family="MIR"
repeat_region 29303..29383
/rpt_family="Alu"
repeat_region 31059..31158
/rpt_family="L1"
repeat_region 31180..31624
/rpt_family="L1"
repeat_region 31626..31773
/rpt_family="Alu"
repeat_region 31774..31993
/rpt_family="Alu"

Query Match 90.1%; Score 1258; DB 9; Length 128466;
Best Local Similarity 99.9%; Pred. No. 0;
Matches 1108; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 62 GAGTACCATGATATCCAGACGACACCATCCCGGAGAACTACCGCTGCGCATCC 121
DB 113798 GAGTACCATGATATCCAGACGACACCATCCCGGAGAACTACCGCTGCGCATCC 113857
QY 122 TACCAACACGAGGAGCTGCTCTCTTCACTGTTCACTGCTGAGGCTGTGATCTGT 181
DB 113858 TACCAACACGAGGAGCTGCTCTCTTCACTGTTCACTGCTGAGGCTGTGATCTGT 113917
QY 182 GAGAGCATGCCCATGCTGCGGCTTTGTGTGACCAACCAACCACTGACAGTAG 241
DB 113918 GAGAGCATGCCCATGCTGCGGCTTTGTGTGACCAACCAACCACTGACAGTAG 113977
QY 242 CCAAGTGGAGAGAGCCCTTCCAGAGAGATGAGAGACCTCTGTGAGGTATAGTAGT 301
DB 113978 CCAAGTGGAGAGAGCCCTTCCAGAGAGATGAGAGACCTCTGTGAGGTATAGTAGT 114037
QY 302 GATCCCCATCGAAAGTCAGAGGGGCTGTGAGGTGATGAGAGAGTATACGTCTT 361
DB 114038 GATCCCCATCGAAAGTCAGAGGGGCTGTGAGGTGATGAGAGAGTATACGTCTT 114097
QY 362 CAAGGAGTCAAATTAAGGAGAAATGCTTGTGCTCCAGAAAGAGAAATCCAGCCCTGT 421
DB 114098 CAAGGAGTCAAATTAAGGAGAAATGCTTGTGCTCCAGAAAGAGAAATCCAGCCCTGT 114157
QY 422 TACCTCTACCTCTGCCCCCAGAGTCGAGCTGCTCTTTTCAAGACTGATGAGCCA 481
DB 114158 TACCTCTACCTCTGCCCCCAGAGTCGAGCTGCTCTTTTCAAGACTGATGAGCCA 114217
QY 482 AGTGTCTCTGATCCCAACAACAACATATGTGAAGGCTTGGCTGACCTATCTGAG 541
DB 114218 AGTGTCTCTGATCCCAACAACAACATATGTGAAGGCTTGGCTGACCTATCTGAG 114277
QY 542 GCTCGGCTACCAAGCTGATATCTTCCAGAGCTGGGCTTGGCTGAGAGGAGTACTTG 601
DB 114278 GCTCGGCTACCAAGCTGATATCTTCCAGAGCTGGGCTTGGCTGAGAGGAGTACTTG 114337
QY 602 CACTGGCAGACATGATGCTACCTGGAGACCCCTGAGACAAAGTAAATCATCCACAG 661
DB 114338 CACTGGCAGACATGATGCTACCTGGAGACCCCTGAGACAAAGTAAATCATCCACAG 114397
QY 662 ACAGATGTGACCAAGCAAAAGTGAATATGCAATATGTAATGTGATTTACGAGC 721
DB 114397 ACAGATGTGACCAAGCAAAAGTGAATATGCAATATGTAATGTGATTTACGAGC 114457
```



Db 114398 ACAGATGTGACGAGCAACGTCGCAATATATGCCAAATGTAAATGTGAGTTTACCAGC 114457  
QY 722 CTAGCTATGGAGTGTGCTGCTCTAGTCCAGAAATCATGGGGGTATGATGCTCTCCAA 781  
Db 114458 CTAGCTATGGAGTGTGCTGCTCTAGTCCAGAAATCATGGGGGTATGATGCTCTCCAA 114517  
QY 782 CCTGTGGGCTGTAGCAAGCTCAGGCTAGTCTCCCACTGGGGGCTGTGCTCTCCG 841  
Db 114518 CCTGTGGGCTGTAGCAAGCTCAGGCTAGTCTCCCACTGGGGGCTGTGCTCTCCG 114577  
QY 842 GAGCGGTTCCGTGGGAGGCGCCCTCACTGTCTCAATAGTGTAGAAATGACTAAAGC 901  
Db 114578 GAGCGGTTCCGTGGGAGGCGCCCTCACTGTCTCAATAGTGTAGAAATGACTAAAGC 114637  
QY 902 CCTGTGCTGCTCTCTCAGTCCAGACAGAGCGGCTGCGGGGCTGCTGGGGCAATCCAT 961  
Db 114638 CCTGTGCTGCTCTCTCAGTCCAGACAGAGCGGCTGCGGGGCTGCTGGGGCAATCCAT 114697  
QY 962 CGTGAAGTGTCTCTCAGCTTAGTCTGACAGAGACTTGGCGGGGATGCTCCAGAT 1021  
Db 114698 CGTGAAGTGTCTCTCAGCTTAGTCTGACAGAGACTTGGCGGGGATGCTCCAGAT 114757  
QY 1022 GTGGGATGTTCTTACTCTGGGGAGGCTATCTTGACTCTCCGACAGGGGACATCCAG 1081  
Db 114758 GTGGGATGTTCTTACTCTGGGGAGGCTATCTTGACTCTCCGACAGGGGACATCCAG 114817  
QY 1082 CCAGCCAGGGGCTCAGGGGAGAGGTGACACCTCAGCATAGACCAAGACTGGGGTCAAG 1141  
Db 114818 CCAGCCAGGGGCTCAGGGGAGAGGTGACACCTCAGCATAGACCAAGACTGGGGTCAAG 114877  
QY 1142 GAGCAGGTGTGTTTATGACAGAGACTTGGGGGCTGGGGGCTTTCTGCTC 1201  
Db 114878 GAGCAGGTGTGTTTATGACAGAGACTTGGGGGCTGGGGGCTTTCTGCTC 114937  
QY 1202 ATTGCTTTCAATGAAGAGCTCAAGAGACCAAGAGCTTTCCCTCTCTCAAGT 1261  
Db 114938 ATTGCTTTCAATGAAGAGCTCAAGAGACCAAGAGCTTTCCCTCTCTCAAGT 114997  
QY 1262 TGAATATCCAGATCTTTTGTACTCTGTGTGTTAATGTTATTTTGTAAATAA 1321  
Db 114998 TGAATATCCAGATCTTTTGTACTCTGTGTGTTAATGTTATTTTGTAAATAA 115057  
QY 1322 AAATAAATATGTTAATAAATGATGTTTACAGCAAACTCTTCCCTAA 1370  
Db 115058 AAATAAATATGTTAATAAATGATGTTTACAGCAAACTCTTCCCTAA 115106

RESULT 3  
AX056460 2461 bp DNA linear PAT 13-JAN-2001  
LOCUS AX056460 Sequence 104 from Patent WO0073469.  
DEFINITION AX056460  
ACCESSION AX056460.1 GI:12229167  
VERSION  
KEYWORDS  
SOURCE human.  
ORGANISM Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.  
REFERENCE  
AUTHORS Ploeman, G.D., Martinez, R., Whyte, D. and Sudersanam, S.  
TITLE Protein kinases  
JOURNAL Patent: WO 0073469-A 104 07-DEC-2000;  
Sugen, Inc. (US)  
FEATURES  
Location/Qualifiers  
1..2461  
/organism="Homo sapiens"  
/db\_xref="taxon:9606"

BASE COUNT 426 a 771 c 811 g 453 t  
ORIGIN  
Query Match 62.7%; Score 876; DB 6; Length 2461;  
Best Local Similarity 99.9%; Pred. No. 0;  
Matches 926; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 442 CAGGTGCGAGCTGTGCTTTTCAAGACTGATGAGAGCCAGTGTCTCCGTATCCCA 501  
Db 1535 CAGGTGCGAGCTGTGCTTTTCAAGACTGATGAGAGCCAGTGTCTCCGTATCCCA 1594  
QY 502 AGACCAATATGTGAAGGCTCTGCTGACCTATCTGAGGGCTCGGCTGACAGTGA 561  
Db 1595 AGACCAATATGTGAAGGCTCTGCTGACCTATCTGAGGGCTCGGCTGACAGTGA 1654  
QY 562 ATCTCTGAGAGCTGGGCTTGCCTGTGAGAGAGTGTGCTGCACTGGCAGACATGATG 621  
Db 1655 ATCTCTGAGAGCTGGGCTTGCCTGTGAGAGAGTGTGCTGCACTGGCAGACATGATG 1714  
QY 622 ACCTGGGAACCCCTGACAGCAAAAGCTTAACATCCAGACAGACAGATGTGACAGCA 681  
Db 1715 ACCTGGGAACCCCTGACAGCAAAAGCTTAACATCCAGACAGACAGATGTGACAGCA 1774  
QY 682 CGTGCATATATGCCAAATGTTAAATGTGAGTTTACAGCTTACTATGAGGACTGCTG 741  
Db 1775 CGTGCATATATGCCAAATGTTAAATGTGAGTTTACAGCTTACTATGAGGACTGCTG 1834  
QY 742 TCTAGTCCAGAAATCATGGGGGTATGACTGCTCCCAACCTGTGGGCTGTAAACAG 801  
Db 1835 TCTAGTCCAGAAATCATGGGGGTATGACTGCTCCCAACCTGTGGGCTGTAAACAG 1894  
QY 802 CTCAGGCTAGTCTCCCACTGGGGGCTGTGCTCTCCCTGAGAGCGGTTCCGTTGGAGC 861  
Db 1895 CTCAGGCTAGTCTCCCACTGGGGGCTGTGCTCTCCCTGAGAGCGGTTCCGTTGGAGC 1954  
QY 862 CCATCACTGTGTTCAATAGTGTGAGATGAGTAAAGCCCTGCTGCTGCTGCA 921  
Db 1955 CCATCACTGTGTTCAATAGTGTGAGATGAGTAAAGCCCTGCTGCTGCTGCA 2014  
QY 922 TGCACAGAGCGGCTGGGGGCTGCTGGGGCAATCCATGCGAGTGTCTCAGT 981  
Db 2015 TGCACAGAGCGGCTGGGGGCTGCTGGGGCAATCCATGCGAGTGTCTCAGT 2074  
QY 982 TAGGTCTGACAGAGACTTGGCGGGGATGCTCCAGATGTTGGTGTATCTGACTG 1041  
Db 2075 TAGGTCTGACAGAGACTTGGCGGGGATGCTCCAGATGTTGGTGTATCTGACTG 2134  
QY 1042 GAGGCTATCTGTGACTCTCCGACAGGGACACTCCAGGGCCAGCCAGGGGTCAAGG 1101  
Db 2135 GAGGCTATCTGTGACTCTCCGACAGGGACACTCCAGGGCCAGCCAGGGGTCAAGG 2194  
QY 1102 AGAGGTGACACTCTGACATGAGCCAAAGCTGGGGTCAAGGACAGTGTGTTAGCC 1161  
Db 2195 AGAGGTGACACTCTGACATGAGCCAAAGCTGGGGTCAAGGACAGTGTGTTAGCC 2254  
QY 1162 AGGACCTGGGGGCGGGGCTGGGGGCTTCTGCTCATTTGCTTTCAATGAAGAGC 1221  
Db 2255 AGGACCTGGGGGCGGGGCTGGGGGCTTCTGCTCATTTGCTTTCAATGAAGAGC 2314  
QY 1222 TCAAGACAGCAAAACAGGCTTTCCCTCTCTCAGTTGATATCCAGAACTTTTG 1281  
Db 2315 TCAAGACAGCAAAACAGGCTTTCCCTCTCTCAGTTGATATCCAGAACTTTTG 2374  
QY 1282 TACTTCTTGTGTTAATGTTTATTTTGTAAATAAATTAATTAATTAATTAATA 1341  
Db 2375 TACTTCTTGTGTTAATGTTTATTTTGTAAATAAATAAATAAATAATTAATTAATA 2434  
QY 1342 ATGATGTTTCAAGCAAACTCTTCCCT 1368  
Db 2435 ATGATGTTTCAAGCAAACTCTTCCCT 2461

RESULT 4  
BC007901 1493 bp mRNA linear PRI 12-JUL-2001  
LOCUS BC007901 Homo sapiens, clone IMAGE:4139786, mRNA, partial cds.  
DEFINITION BC007901  
ACCESSION BC007901.1 GI:14043927  
VERSION  
KEYWORDS



SOURCE	Homo sapiens.
ORGANISM	Homo sapiens.
	Eulaxyota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi
	Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE	1 (bases 1 to 1493)
AUTHORS	Strausberg,R.
TITLE	Direct Submission
JOURNAL	Submitted (11-MY-2001) National Institutes of Health, Mammalian

BASE COUNT	98 a	/db_xref="taxon:9606"	122 c	147 g	103 t
Query Match	31.6%	Score 442;	DB 6;	Length 470;	
Best Local Similarity	100.0%;	Pred. No. 4,2e-249;			
Matches 442;	Conservative 0;	Mismatches 0;	Indels 0;	Gaps 0;	
OY	572	GCTGGGCTTGCTGTGAGAGAGTGACTTGCACTGGCAGACTGCATGTCACTGGGAAAC	631		
DB	7	GCTGGGCTTGCTGTGAGAGAGTGACTTGCACTGGCAGACTGCATGTCACTGGGAAAC	66		
OY	632	CCCTGCAGCAAAAGCTTAACATCCAGACAGACAGATGTGACAGGACAAAGTGCATTA	691		
DB	67	CCCTGCAGCAAAAGCTTAACATCCAGACAGACAGATGTGACAGGACAAAGTGCATTA	126		
OY	692	TGCCAAATGTTAAATGTGAGTTTACCAAGCTAGTATGGAGCTGCTGGCTCTTATGTC	751		
DB	127	TGCCAAATGTTAAATGTGAGTTTACCAAGCTAGTATGGAGCTGCTGGCTCTTATGTC	186		
OY	752	GGAATCATGAGGGGATATGACTGCTCTTCCAACTCTGGGCTGTAAAGCAAGCTCAGGCT	811		
DB	187	GGAATCATGAGGGGATATGACTGCTCTTCCAACTCTGGGCTGTAAAGCAAGCTCAGGCT	246		
OY	812	TCTCCCACTGGGGGCTGTGACCCCTCCCTGCGGAGACGGTTCCGGGGAGGCCCATCACT	871		
DB	247	TCTCCCACTGGGGGCTGTGACCCCTCCCTGCGGAGACGGTTCCGGGGAGGCCCATCACT	306		
OY	872	GTTCAATAGTGTGAGATATGTAGCTTAAGCCCTGTGCTGTCTGTCTGACATGCCACACA	931		
DB	307	GTTCAATAGTGTGAGATATGTAGCTTAAGCCCTGTGCTGTCTGTCTGACATGCCACACA	366		
OY	932	GGCGGTGGGGGCTGGCTGGGGGCAATCCATCTGTGAGTCTTCTCAGCTTAAGTCTGGA	991		
DB	367	GGCGGTGGGGGCTGGCTGGGGGCAATCCATCTGTGAGTCTTCTCAGCTTAAGTCTGGA	426		
OY	992	CAGGAGACTTGCGCGGGGATGC 1013			
DB	427	CAGGAGACTTGCGCGGGGATGC 448			
RESULT 6					
AX408491/c		427 bp	DNA	linear	PAT 14-JUN-2002
LOCUS	AX408491				
DEFINITION	Sequence 1138 from Patent WO0229103.				
ACCESSION	AX408491				
VERSION	AX408491.1	GI:21441196			
KEYWORDS	human.				
SOURCE	Homo sapiens				
ORGANISM	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;				
REFERENCE	Mammalia; Euteleostomi; Primates; Catarrhini; Homiidae; Homo.				
AUTHORS	1				
TITLE	Alvares, C., Horne, D., Peres-da-Silva, S. and Vockley, J.G.				
JOURNAL	Gene expression profiles in liver cancer				
GENE LOGIC	Patent: WO 0229103-A 1138 11-APR-2002;				
FEATURES	GENE LOGIC INC (US)				
source	Location/Qualifiers				
BASE COUNT	107 a 126 c 92 g 102 t				
ORIGIN					
Query Match	23.0%;	Score 322;	DB 6;	Length 427;	
Best Local Similarity	99.5%;	Pred. No. 3e-178;			
Matches 422;	Conservative 0;	Mismatches 2;	Indels 0;	Gaps 0;	
OY	936	GTGGGGGCTGGCTGGGGAACAATCCATCGTGGAGCTTCTCAGCTTAAGTCTGACAG	995		
DB	427	GTGGGGGCTGGCTGGGGAACAATCCATCGTGGAGCTTCTCAGCTTAAGTCTGACAG	368		

QY		996	AGACTGGCGGGGATGCTCCAGGATGCGGTATTTCGTACCTGGGGAGGCTATCTTG	1055
Db		367	AGACTGGCGGGGATGCTCCAGGATGCGGTATTTCGTACCTGGGGAGGCTATCTTG	308
QY		1056	ACCTCCGACAGGGGACACTCCAGGCAGGCCAGGGGTCAGGGGACAGAGTGCACA	1115
Db		307	ACCTCCGACAGGGGACACTCCAGGCAGGCCAGGGGTCAGGGGACAGAGTGCACA	248
QY		1116	CAGCATGAGCCAAAGACTGCGGGTCAGGGAGCAGGTGTGGTTTGAGCCAGAGCCTG	1175
Db		247	CAGCATGAGCCAAAGACTGCGGGTCAGGGAGCAGGTGTGGTTTGAGCCAGAGCCTG	188
QY		1176	GCGTGGGGCCCGGGGCTTTCTGCTCAATTTGCTTTCATGAAGAACCCTTAAGC	1235
Db		187	GCGTGGGGCCCGGGGCTTTCTGCTCAATTTGCTTTCATGAAGAACCCTTAAGC	128
QY		1236	ACGAGGCTTTCCCCCTTCCTGAGTTGAATATCCAGAACTTTTGACTCTTGTTGT	1295
Db		127	ACGAGGCTTTCCCCCTTCCTGAGTTGAATATCCAGAACTTTTGACTCTTGTTGT	68
QY		1296	TAAATGTTTTATTTTGTAAAAAATAAAAATTAATTAATTAATTAATTAATTAAT	1355
Db		67	TAAATGTTTTATTTTGTAAAAAATAAAAATTAATTAATTAATTAATTAATTAAT	8
QY		1356	CMAA 1359	
Db		7	CMAA 4	
RESULT 7				
LOCUS		G21939	462 bp	DNA linear STS 31-MAY-1996
DEFINITION		G21939	human STS WI-15430, sequence tagged site.	
ACCESSION		G21939		
VERSION		G21939.1	GI:1342265	
KEYWORDS		STS; STS sequence; primer; sequence tagged site.		
SOURCE		Homo sapiens STS derived from sequences in dbEST and the Unigene collection.		
ORGANISM		Homo sapiens		
REFERENCE		Mammalia; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;		
AUTHORS		Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.		
TITLE		1 (bases 1 to 462)		
JOURNAL		Hudson, T.		
COMMENT		Whitehead Institute/MIT Center for Genome Research; Physically Mapped STS		
		Unpublished (1995)		
		Contact: Thomas Hudson		
		Whitehead Institute/MIT Center for Genome Research		
		Whitehead Institute for Biomedical Research		
		9 Cambridge Center, Cambridge MA 02142 USA		
		Tel: 617 252 1900		
		Fax: 617 252 1902		
		Email: thudson@genome.wi.mit.edu		
		Primer A: AGGAGAAGTTTCTGTGAA		
		Primer B: GCCTCAAAGCAGCCAAC		
		STS size: 150		
		PCR Profile:		
		Denaturation:		
		Annealing: 56 degrees C		
		Polymerization:		
		PCR Cycles: 35		
		Thermal Cycler:		
		Protocol:		
		Template: 10 ng		
		Primer: each 5 pM		
		dNTPs: each 4 mM		
		Taq Polymerase: 0.025 units/ul		
		Total Vol.: 20 ul		
		Buffer:		



```

* 9745 9844: gap of unknown length
* 9845 11023: contig of 1179 bp in length
* 11024 11123: gap of unknown length
* 11124 12519: contig of 1396 bp in length
* 12520 12619: gap of unknown length
* 12620 13673: contig of 1054 bp in length
* 13674 13773: gap of unknown length
* 13774 15620: contig of 1847 bp in length
* 15621 15720: gap of unknown length
* 15721 17466: contig of 1746 bp in length
* 17467 17566: gap of unknown length
* 17567 19136: contig of 1570 bp in length
* 19137 19236: gap of unknown length
* 19237 20973: contig of 1737 bp in length
* 20974 21073: gap of unknown length
* 21074 22091: contig of 1018 bp in length
* 22092 22191: gap of unknown length
* 22192 23778: contig of 1587 bp in length
* 23779 23878: gap of unknown length
* 23879 25293: contig of 1415 bp in length
* 25294 25983: gap of unknown length
* 25984 26983: contig of 1590 bp in length
* 26984 27083: gap of unknown length
* 27084 29408: contig of 2325 bp in length
* 29409 29508: gap of unknown length
* 29509 31243: contig of 1735 bp in length
* 31244 31343: gap of unknown length
* 31344 33604: contig of 2261 bp in length
* 33605 33704: gap of unknown length
* 33705 35038: contig of 1334 bp in length
* 35039 35138: gap of unknown length
* 35139 36399: contig of 1261 bp in length
* 36400 37623: gap of unknown length
* 37624 37723: gap of unknown length
* 37724 40109: contig of 2376 bp in length
* 40100 40199: gap of unknown length
* 40200 42674: contig of 2475 bp in length
* 42675 42774: gap of unknown length
* 42775 45617: contig of 2843 bp in length
* 45618 45717: gap of unknown length
* 45718 47425: contig of 1708 bp in length
* 47426 47525: gap of unknown length
* 47526 48809: contig of 1284 bp in length
* 48810 48909: gap of unknown length
* 48910 50692: contig of 1783 bp in length
* 50693 50792: gap of unknown length
* 50793 52171: contig of 1379 bp in length
* 52172 52271: gap of unknown length
* 52272 53845: contig of 1574 bp in length
* 53846 53945: gap of unknown length
* 53946 56501: contig of 2556 bp in length
* 56502 56601: gap of unknown length
* 56602 59622: contig of 3021 bp in length
* 59623 59723: gap of unknown length
* 59724 62615: contig of 2893 bp in length
* 62616 62715: gap of unknown length
* 62716 65623: contig of 2908 bp in length
* 65624 65723: gap of unknown length
* 65724 67513: contig of 1790 bp in length
* 67514 70311: contig of 2698 bp in length
* 70312 70411: gap of unknown length
* 70412 72769: contig of 2358 bp in length
* 72770 72869: gap of unknown length
* 72870 76193: contig of 3324 bp in length
* 76194 76293: gap of unknown length
* 76294 79434: contig of 3141 bp in length
* 79435 79534: gap of unknown length
* 79535 84177: contig of 4643 bp in length
* 84178 84277: gap of unknown length
* 84278 88578: contig of 4301 bp in length
* 88579 88678: gap of unknown length

```

```

FEATURES             * 88679 94366: contig of 5688 bp in length.
source               1..94366
                    /organism="Rattus norvegicus"
                    /db_xref="taxon:10116"
                    /clone="CH250-12986"
BASE COUNT          22805 a 22369 c 21709 g 21135 t 4348 others
ORIGIN
Query Match          3.1% Score 43; DB 2; Length 94366;
Best Local Similarity 100.0%; Pred. No. 1.8e-13;
Matches 43; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 203 GCCTTGTGTCACACACACACACCTTGAGAGGTGACCCAG 245
Db 24116 GCCTTGTGTCACACACACACACCTTGAGAGGTGACCCAG 24158

RESULT 9
AC122406/c          176771 bp DNA linear HTG 20-JUN-2002
LOCUS               Mus musculus chromosome UNK clone RP24-126L14, WORKING DRAFT
DEFINITION          SEQUENCE, 12 unordered pieces.
ACCESSION           AC122406
VERSION             AC122406.2 GI:21490547
KEYWORDS            HTG; HTGS PHASE1; HTGS_DRAFT; HTGS_FULLTOP.
SOURCE              house mouse.
ORGANISM            Mus musculus
                    Mammalia; Euteleostomi;
                    Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
                    Mammalia; Euteleostomi; Rodentia; Sciurognathi; Muridae; Mus.
REFERENCE
AUTHORS             McPherson,J.D. and Waterston,R.H.
TITLE               The sequence of Mus musculus clone
JOURNAL             Unpublished
REFERENCE
AUTHORS             McPherson,J.D. and Waterston,R.H.
TITLE               Direct Submissio
JOURNAL             Submitted (23-MAY-2002) Genome Sequencing Center, 4444 Forest Park
                    Parkway, St. Louis, MO 63108, USA
                    3 (bases 1 to 176771)
REFERENCE
AUTHORS             McPherson,J.D. and Waterston,R.H.
TITLE               Direct Submissio
JOURNAL             Submitted (20-JUN-2002) Genome Sequencing Center, 4444 Forest Park
                    Parkway, St. Louis, MO 63108, USA
                    On Jun 20, 2002 this sequence version replaced gi:21105866.
COMMENT
----- Genome Center -----
Center: Washington University Genome Sequencing Center
Center code: WUGSC
Web site: http://genome.wustl.edu/gsc/index.shtml
Contact: submissions@watson.wustl.edu
----- Project Information -----
Center project name: M_BB0126L14
----- Summary Statistics -----
Sequencing vector: M13; 0%
Sequencing vector: plasmid; 100%
Chemistry: Dye-terminator Big Dye; 100% of reads
Chemistry: Dye-terminator Big Dye; 100% of reads
Assembly program: Phrap; version 0.990319
Consensus quality: 171598 bases at least Q40
Consensus quality: 172211 bases at least Q30
Consensus quality: 172737 bases at least Q20
Insert size: 186000; agarose-tp
Insert size: 175614; sum-of-ctnigs
Quality coverage: 11.87 in Q20 bases; agarose-tp
Quality coverage: 11.02 in Q20 bases; sum-of-ctnigs
-----
* NOTE: This is a 'working draft' sequence. It currently
* consists of 12 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.

```

\* This record will be updated with the finished sequence  
\* as soon as it is available and the accession number will  
\* be preserved

1 57: contig of 57 bp in length  
58 157: gap of unknown length  
157 3050: contig of 2893 bp in length  
3051 3150: gap of unknown length  
3151 5182: contig of 2032 bp in length  
5183 5283: gap of unknown length  
5283 14170: contig of 8888 bp in length  
14171 14270: gap of unknown length  
14271 22115: contig of 7845 bp in length  
22116 32754: gap of unknown length  
32755 32854: contig of 10539 bp in length  
32855 46558: gap of unknown length  
46559 46659: gap of unknown length  
46660 65170: contig of 18512 bp in length  
65171 65270: gap of unknown length  
65271 83892: contig of 18622 bp in length  
83893 83992: gap of unknown length  
83993 103925: contig of 19933 bp in length  
103926 104025: gap of unknown length  
104026 133284: contig of 29259 bp in length  
133285 133384: gap of unknown length  
133385 176771: contig of 43387 bp in length.

location/Qualifiers  
1. 176771  
/organism="Mus musculus"  
/db\_xref="taxon:10090"  
/chromosome="X"

misc\_feature 1 57  
/note="assembly\_name:Contig11"  
misc\_feature 158 3050  
/note="assembly\_name:Contig15"  
misc\_feature 3151 5182  
/note="assembly\_name:Contig16"  
misc\_feature 5283 14170  
/note="assembly\_name:Contig17"  
misc\_feature 14271 22115  
/note="assembly\_name:Contig18"  
misc\_feature 22116 32754  
/note="assembly\_name:Contig19"  
misc\_feature 32855 46558  
/note="assembly\_name:Contig20"  
misc\_feature 46559 46659  
/note="assembly\_name:Contig21"  
misc\_feature 46660 65170  
/note="assembly\_name:Contig22"  
misc\_feature 65171 65270  
/note="assembly\_name:Contig23"  
misc\_feature 65271 83892  
/note="assembly\_name:Contig24"  
misc\_feature 83893 83992  
/note="assembly\_name:Contig25"  
misc\_feature 103926 104025  
/note="assembly\_name:Contig26"  
misc\_feature 104026 133284  
/note="assembly\_name:Contig27"  
misc\_feature 133285 133384  
/note="assembly\_name:Contig28"  
misc\_feature 133385 176771  
/note="assembly\_name:Contig29"

BASE COUNT 45173 a 42065 c 42632 g 45785 t 1116 others

ORIGIN

Query Match 3.1%; Score 43; DB 2; Length 176771;  
Best Local Similarity 100.0%; Pred. No. 1.8e-13;  
Matches 43; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 203 GCCTTTGTGTCACCAACGACGACCTGAGCGGTGAGCCAG 245  
|||||  
DB 57080 GCCTTTGTGTCACCAACGACGACCTGAGCGGTGAGCCAG 57038  
|||||

RESULT 10  
BC022157 1586 bp mRNA linear ROD 07-AUG-2002  
LOCUS BC022157  
DEFINITION Mus musculus, clone IMAGE:5134400, mRNA, partial cds.  
ACCESSION BC022157

VERSION BC022157.1 GI:18381057  
KEYWORDS house mouse.  
SOURCE Mus musculus.  
ORGANISM Mus musculus.  
REFERENCE 1 (bases 1 to 1586)  
AUTHORS Strausberg, R.  
TITLE Direct Submission  
JOURNAL Submitted (24-JAN-2002) National Institutes of Health, Mammalian Gene Collection (MGC), Cancer Genomics Office, National Cancer Institute, 31 Center Drive, Room 11A03, Bethesda, MD 20892-2590, USA

REMARK NIH-MGC Project URL: <http://mgc.ncl.nih.gov>.  
COMMENT Contact: MGC help desk  
Email: [cgabbs-r@mail.nih.gov](mailto:cgabbs-r@mail.nih.gov)  
Tissue Procurement: Jeffrey E. Green, M.D.  
CDNA Library Preparation: Life Technologies, Inc.  
CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LNL)  
DNA Sequencing by: Genome Sequence Centre,  
BC Cancer Agency, Vancouver, BC, Canada  
[info@cgsc.bc.ca](mailto:info@cgsc.bc.ca)  
Steven Jones, Jennifer Asano, Ian Bosdet, Yaron Butterfield,  
Susanna Chan, Readman Chiu, Chris Fjell, Erin Garland, Ran Guin,  
Leticia Hsiao, Martin Krzywicki, Reta Kutche, Oliver Lee, Soo  
Sen Lee, Victor Ling, Carrie Mathewson, Candice McLeavy, Steven  
Ness, Pawan Pandon, Anna-Lisa Prabhu, Parvaneh Saeedi, Jacqueline  
Schein, Duane Smallus, Michael Smith, Lorraine Spence, Jeff Stott,  
Michael Thorne, Miranada Teal, Natasja van den Bosch, Jill Vardy,  
George Yang, Scott Zuyderduyn, Marco Marra.

Clone distribution: MGC clone distribution information can be found  
through the I.M.A.G.E. Consortium/LNL at: <http://image.lnl.gov>  
Series: IRAX Plates: 57 Row: m Column: 21  
This clone was selected for full length sequencing because it  
passed the following selection criteria: Hexamer frequency ORF  
analysis.

FEATURES  
source 1. 1586  
/organism="Mus musculus"  
/db\_xref="taxon:10090"  
/map="FVB/N"  
/clone="IMAGE:5134400"  
/issue="type="Liver, normal, 5 month old male mouse."  
/clone\_lib="NCI CGAP\_L19"  
/lab\_host="DH10B"  
/note="Vector: pCMV-SPORT6"  
CDS  
1. 809  
/codon\_start=3  
/product="unknown (protein for IMAGE:5134400)"  
/protein\_id="AAH22157.1"  
/db\_xref="GI:18381058"  
translation="IPDTLTITTELCAPVEMITQLQTSWEDFRICLSLGRLLHLAH  
SLIGSVTLDEPRPFLVNGELKATDDADVEETPCTSSADCTLEPPARVSLPCS  
AQCCEGMEKERNLYNAYRPFETVLPSPAPSLRLDLSINATGELAMGVDELTAO  
LETAHLPRSGQYQNTSSRAVORIPDSATIGODRYCWSBYHGGCLISFNLALAA  
IDVESHQCAFAVNTQNTMTGRKLVFFKIGMNOVDPAGKTTVXAPG"

BASE COUNT 354 a 455 c 441 g 336 t

ORIGIN

Query Match 2.6%; Score 36; DB 10; Length 1586;  
Best Local Similarity 100.0%; Pred. No. 2.3e-09;  
Matches 36; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 203 GCCTTTGTGTCACCAACGACGACCTGAGCGGT 238  
|||||  
DB 690 GCCTTTGTGTCACCAACGACGACCTGAGCGGT 725  
|||||

RESULT 11  
AX056461 1886 bp DNA linear PAT 13-JAN-2001  
LOCUS AX056461  
DEFINITION Sequence 105 from Patent WO0073469.

ACCESSION AX056461.1 GI:12229168  
 VERSION AX056461.1  
 KEYWORDS  
 SOURCE  
 ORGANISM  
 Murinae gen. sp.  
 Murinae gen. sp.  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae.  
 REFERENCE  
 1 (bases 1 to 1886)  
 Plozman, G.D., Martinez, R., Whyte, D. and Sudersanam, S.  
 TITLE  
 Protein kinases  
 JOURNAL  
 Patent: WO 0073469-A 105 07-DEC-2000;  
 Sugen, Inc. (US)  
 FEATURES  
 source  
 1.1886  
 Location/Qualifiers  
 /organism="Murinae gen. sp."  
 /db\_xref="taxon:39108"  
 BASE COUNT 384 a 559 c 551 g 390 t  
 ORIGIN  
 Query Match 2.6%; Score 36; DB 6; Length 1886;  
 Best Local Similarity 100.0%; Pred. No. 2.3e-09;  
 Matches 36; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 Oy 203 GCCTTGTGCTACCAACCAAGACCTGACAGGT 238  
 Db 981 GCCTTGTGCTACCAACCAAGACCTGACAGGT 1016  
 RESULT 12  
 AC11157/c  
 LOCUS  
 DEFINITION  
 Homo sapiens chromosome 8 clone RP11-70C11 map 8, LOW-PASS SEQUENCE  
 SAMPLING.  
 AC11157  
 Homo sapiens chromosome 8 clone RP11-70C11 map 8, LOW-PASS SEQUENCE  
 AC11157.1 GI:18699922  
 HTG; HTGS\_PHASE0.  
 VERSION  
 HTG; HTGS\_PHASE0.  
 SOURCE  
 Homo sapiens  
 ORGANISM  
 Homo sapiens  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.  
 REFERENCE  
 1 (bases 1 to 69492)  
 Birren, B., Linton, L., Nusbaum, C. and Lander, E.  
 JOURNAL  
 Homo sapiens chromosome 8, clone RP11-70C11  
 REFERENCE  
 2 (bases 1 to 69492)  
 Birren, B., Linton, L., Nusbaum, C., Lander, E., Ali, A., Allen, N.,  
 Anderson, S., Barna, N., Bastien, V., Boguslavsky, L., Boukhgalter, B.,  
 Brown, A., Camarata, J., Campoliano, A., Chang, J., Chazaro, B.,  
 Choquet, Y., Colangelo, M., Collins, S., Collymore, A., Cook, A.,  
 Cooke, P., Dearellano, K., Dewar, K., Diaz, J.S., Dodge, S., Faro, S.,  
 Ferreira, P., Fitzhugh, W., Gage, D., Galagan, J., Gardyna, S.,  
 Ginde, S., Gord, S., Goyette, M., Graham, L., Grand-Pierre, N.,  
 Hagos, B., Horton, L., Hulme, W., Iliev, I., Johnson, R., Jones, C.,  
 Kamat, A., Karatas, A., Kells, C., Labocque, K., Lamazares, R.,  
 Landers, T., Lehoczek, J., Levine, R., Liu, G., Maclean, C.,  
 MacDonald, P., Major, J., Margis, N., Matthews, C., McCarthy, M.,  
 McEwan, P., McKernan, K., Meldrum, J., Meneus, L., Mihova, T.,  
 Mlenga, V., Murphy, T., Naylor, J., Nguyen, C., Nicol, R., Notcu, C.,  
 Norman, C.H., O'Connor, T., O'Donnell, P., O'Neill, D., Oliver, J.,  
 Peterson, K., Phunkhang, P., Pierre, N., Pollara, V., Raymond, C.,  
 Retta, R., Rieback, M., Riley, R., Rise, C., Rogov, P., Roman, J.,  
 Rosetti, M., Roy, A., Santos, R., Schauer, S., Schnapack, R., Seaman, S.,  
 Severy, P., Spencer, B., Stange, Thoman, N., Stejanovic, N.,  
 Strube, N., Sudramanian, A., Talamas, J., Testaye, S., Theodore, J.,  
 Topham, K., Travers, M., Travis, N., Trigilio, J., Vassiliev, H.,  
 Viet, R., Vo, A., Wilson, B., Wu, X., Wyman, D., Ye, W. J., Young, G.,  
 Zainoun, J., Zembek, L., Zimmer, A. and Zody, M.  
 TITLE  
 Direct Submissions  
 JOURNAL  
 Submitted (18-FEB-2002) Whitehead Institute/MIT Center for Genome  
 Research, 320 Charles Street, Cambridge, MA 02141, USA  
 COMMENT  
 All repeats were identified using RepeatMasker:  
 Smit, A.F.A. & Green, P. (1996-1997)  
 http://ftp.genome.washington.edu/RM/RepeatMasker.html

----- Genome Center  
 Center: Whitehead Institute/ MIT Center for Genome Research  
 Center code: MIBR  
 Web site: http://www-seq.wi.mit.edu  
 Contact: sequence\_submissions@genome.wi.mit.edu  
 Project Information  
 Center project name: L25396  
 Center clone name: 70\_C\_11  
 -----  
 \* NOTE: This record contains 86 individual  
 \* sequencing reads that have not been assembled into  
 \* contigs. Runs of N are used to separate the reads  
 \* and the order in which they appear is completely  
 \* arbitrary. Low-pass sequence sampling is useful for  
 \* identifying clones that may be gene-rich and allows  
 \* overlap relationships among clones to be deduced.  
 \* However, it should not be assumed that this clone  
 \* will be sequenced to completion. In the event that  
 \* the record is updated, the accession number will  
 \* be preserved.  
 1  
 706 805: gap of 100 bp  
 806 1512: contig of 707 bp in length  
 1513 1612: gap of 100 bp  
 1613 2325: contig of 713 bp in length  
 2326 2425: gap of 100 bp  
 2426 3121: contig of 696 bp in length  
 3122 3221: gap of 100 bp  
 3222 3930: contig of 709 bp in length  
 3931 4030: gap of 100 bp  
 4031 4744: contig of 714 bp in length  
 4745 4844: gap of 100 bp  
 4845 5561: contig of 717 bp in length  
 5562 5661: gap of 100 bp  
 5662 6390: contig of 729 bp in length  
 6391 6490: gap of 100 bp  
 6491 7201: contig of 711 bp in length  
 7202 7301: gap of 100 bp  
 7302 8011: contig of 710 bp in length  
 8012 8111: gap of 100 bp  
 8112 8829: contig of 718 bp in length  
 8830 8929: gap of 100 bp  
 8930 9616: contig of 687 bp in length  
 9617 9716: gap of 100 bp  
 9717 10409: contig of 693 bp in length  
 10410 10509: gap of 100 bp  
 10510 11220: contig of 711 bp in length  
 11221 11320: gap of 100 bp  
 11321 12020: contig of 700 bp in length  
 12021 12120: gap of 100 bp  
 12121 12842: contig of 722 bp in length  
 12843 12942: gap of 100 bp  
 12943 13641: contig of 699 bp in length  
 13642 13741: gap of 100 bp  
 13742 14470: contig of 729 bp in length  
 14471 14570: gap of 100 bp  
 14571 15280: contig of 710 bp in length  
 15281 15380: gap of 100 bp  
 15381 16097: contig of 717 bp in length  
 16098 16197: gap of 100 bp  
 16198 16910: contig of 713 bp in length  
 16911 17010: gap of 100 bp  
 17011 17707: contig of 697 bp in length  
 17708 17807: gap of 100 bp  
 17808 18509: contig of 702 bp in length  
 18510 18609: gap of 100 bp  
 18610 19314: contig of 705 bp in length  
 19315 19414: gap of 100 bp  
 19415 20120: contig of 706 bp in length  
 20121 20220: gap of 100 bp  
 20221 20925: contig of 705 bp in length  
 20926 21025: gap of 100 bp  
 21026 21730: contig of 705 bp in length

21731 21830: gap of 100 bp  
21831 22540: contig of 710 bp in length  
22541 22640: gap of 100 bp  
22641 23362: contig of 722 bp in length  
23363 23462: gap of 100 bp  
23463 24186: contig of 724 bp in length  
24187 24286: gap of 100 bp  
24287 25003: contig of 717 bp in length  
25004 25103: gap of 100 bp  
25104 25812: contig of 709 bp in length  
25813 25912: gap of 100 bp  
25913 26629: contig of 717 bp in length  
26630 26729: gap of 100 bp  
26730 27427: contig of 698 bp in length  
27428 27527: gap of 100 bp  
27528 28249: contig of 722 bp in length  
28250 28349: gap of 100 bp  
28350 29065: contig of 716 bp in length  
29066 29165: gap of 100 bp  
29166 29868: contig of 703 bp in length  
29869 29968: gap of 100 bp  
29969 30686: contig of 718 bp in length  
30687 30786: gap of 100 bp  
30787 31496: contig of 710 bp in length  
31497 31596: gap of 100 bp  
31597 32358: contig of 762 bp in length  
32359 32458: gap of 100 bp  
32459 33171: contig of 713 bp in length  
33172 33271: gap of 100 bp  
33272 33982: contig of 711 bp in length  
33983 34082: gap of 100 bp  
34083 34769: contig of 687 bp in length  
34770 34869: gap of 100 bp  
34870 35573: contig of 704 bp in length  
35574 35673: gap of 100 bp  
35674 36382: contig of 709 bp in length  
36383 36482: gap of 100 bp  
36483 37180: contig of 698 bp in length  
37181 37280: gap of 100 bp  
37281 37983: contig of 703 bp in length  
37984 38083: gap of 100 bp  
38084 38799: contig of 716 bp in length  
38800 38899: gap of 100 bp  
38900 39611: contig of 712 bp in length  
39612 39711: gap of 100 bp  
39713 40424: contig of 713 bp in length  
40425 40524: gap of 100 bp  
40525 41242: contig of 718 bp in length  
41243 41342: gap of 100 bp  
41343 42052: contig of 710 bp in length  
42053 42152: gap of 100 bp  
42153 42857: contig of 705 bp in length  
42858 42957: gap of 100 bp  
42959 43670: contig of 713 bp in length  
43671 43770: gap of 100 bp  
43771 44477: contig of 707 bp in length  
44478 44577: gap of 100 bp  
44579 45283: contig of 706 bp in length  
45284 45383: gap of 100 bp  
45384 46067: contig of 684 bp in length  
46068 46167: gap of 100 bp  
46168 46877: contig of 710 bp in length  
46878 46977: gap of 100 bp  
46979 47680: contig of 683 bp in length  
47681 47760: gap of 100 bp  
47761 48476: contig of 716 bp in length  
48477 48576: gap of 100 bp  
48578 49290: contig of 714 bp in length  
49291 49390: gap of 100 bp  
49391 50111: contig of 721 bp in length  
50112 50211: gap of 100 bp  
50212 50913: contig of 702 bp in length  
50914 51013: gap of 100 bp

Query Match 2.4% Score 33; DB 2; Length 69492;  
Best Local Similarity 100.0%; Pred. No. 1.4e-07;  
Matches 33; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
Qy 1359 ACTCTCCCTAAATAAAAAAAAAAAAAAAAAAAG 1391  
Db 67523 ACTCTCCCTAAATAAAAAAAAAAAAAAAAAAAG 67491  
RESULT 13  
AC091069/c  
LOCUS AC091069/c  
DEFINITION Homo sapiens chromosome 3 clone RP11-161l3 map 3p, complete  
ACCESSION AC091069  
VERSION AC091069.1 GI:13447410  
KEYWORDS HTG.  
SOURCE Homo sapiens.  
ORGANISM Homo sapiens  
REFERENCE  
AUTHORS  
1 (bases 1 to 164796)  
Bao, W., Bao, J., Bao, Q., Bian, X., Cao, T., Chen, C., Chen, J., Ding, H.,  
Dong, W., Fan, H., Feng, X., Gong, J., Guan, Q., Gu, X., Guo, D., Guo, Z.,  
He, L., Hu, S., Huang, F., Jin, Y., Kang, N., Li, C., Li, C., Li, F.,  
Li, G., Li, J., Li, L., Li, S., Li, T., Liu, Y., Liu, N., Liu, B., Liu, Y.,  
Li, W., Li, Y., Luo, J., Luo, Y., Qi, O., Qi, X., Song, L.,  
Song, S., Sun, M., Sun, W., Sun, Y., Tan, X., Tao, R., Wang, H., Wang, J.,  
Wang, J., Wang, J., Wang, L., Wang, L., Wang, R., Wang, X.,  
Wang, Y., Wu, D., Wu, Q., Xie, F., Xian, Z., Xue, Y., Yan, C., Yang, X.,  
Yu, B., Zeng, Y., Zhang, G., Zhang, H., Zhang, H., Zhang, L., Zhang, M.,  
Zhang, X., Zhang, X., Zhang, Y., Zhang, Y., Zhang, Z., Zhu, B., Zhu, N.,  
Yu, J. and Yang, H.  
Unpublished  
TITLE  
JOURNAL  
REFERENCE  
AUTHORS  
2 (bases 1 to 164796)  
Bao, W., Bao, J., Bao, Q., Bian, X., Cao, T., Chen, C., Chen, J., Ding, H.,  
Dong, W., Fan, H., Feng, X., Gong, J., Guan, Q., Gu, X., Guo, D., Guo, Z.,  
He, L., Hu, S., Huang, F., Jin, Y., Kang, N., Li, C., Li, C., Li, F.,  
Li, G., Li, J., Li, L., Li, S., Li, T., Liu, Y., Liu, N., Liu, B., Liu, Y.,  
Li, W., Li, Y., Luo, J., Luo, Y., Qi, O., Qi, X., Song, L.,  
Song, S., Sun, M., Sun, W., Sun, Y., Tan, X., Tao, R., Wang, H., Wang, J.,  
Wang, J., Wang, J., Wang, L., Wang, L., Wang, R., Wang, X.,  
Wang, Y., Wu, D., Wu, Q., Xie, F., Xian, Z., Xue, Y., Yan, C., Yang, X.,  
Yu, B., Zeng, Y., Zhang, G., Zhang, H., Zhang, H., Zhang, L., Zhang, M.,  
Zhang, X., Zhang, X., Zhang, Y., Zhang, Y., Zhang, Z., Zhu, B., Zhu, N.,  
Yu, J. and Yang, H.  
Direct Submission  
Submitted (26-MAR-2001) Human Genomic Center, Institute of  
Genetics, Chinese Academy of Sciences, Datun Road, Beijing, Beijing  
100101, P.R.China  
COMMENT  
-----Genome Center  
Center: Beijing Center  
Center code: Beijing  
Website: http://hgsc.iqtp.ac.cn  
http://www.genomics.org.cn  
Contact: hgsc@iqtp.ac.cn  
-----Project Information  
Center project name: 1# project  
Center clone name: Rp11-161l3  
-----Summary Statistics



Sequencing vector: pUC18, 100% of reads  
 Chemistry: Dye-terminator: ET 55% of reads  
 Chemistry: Dye-terminator: Big Dye, 45% of reads  
 Assembly program: Phrap; version 0.990329  
 Consensus quality: 806 bases at least Q40  
 Consensus quality: 1112 bases at least Q30  
 Consensus quality: 1234 bases at least Q20  
 Insert size: 1268; sum-of-contigs  
 Quality coverage: 1.34x in Q20 bases; sum-of-contigs

# FEATURES

## source

Location/Qualifiers  
 1. 164796  
 /organism="Homo sapiens"  
 /db\_xref="taxon:9606"  
 /chromosome="3"  
 /map="3p"  
 /clone="RP11-161L3"

BASE COUNT 54122 a 32035 c 31236 g 47403 t  
 ORIGIN

Query Match 2.3%; Score 32; DB 9; Length 164796;  
 Best Local Similarity 100.0%; Pred. No. 5.7e-07;  
 Matches 32; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1358 AACTCTTCCTATAAAAAAAAAAAAAA 1389  
 DB 100402 AACTCTTCCTATAAAAAAAAAAAAAA 100371

RESULT 14 AC011963 165201 bp DNA linear HTG 21-APR-2000  
 AC011963/c Homo sapiens chromosome 4 clone RP11-173F11 map 4, WORKING DRAFT  
 LOCUS  
 DEFINITION  
 SEQUENCE 10 unordered pieces.

AC011963 GI:7630665  
 HTG; HTGS\_PHASE1; HTGS\_DRAFT.  
 VERSION  
 KEYWORDS  
 SOURCE  
 ORGANISM  
 Homo sapiens  
 Homo sapiens  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.

REFERENCE 1 (bases 1 to 165201)  
 Birren, B., Linton, L., Nuebaum, C., Lander, E., Allen, N., Anderson, M.,  
 Baldwin, J., Batra, N., Beckerly, R., Boguslavsky, L., Boukhgalter, B.,  
 Brown, A., Castle, A., Colangelo, M., Collins, S., Collymore, A.,  
 Cooke, P., Dearellano, K., Dewar, K., Domingo, M., Donelan, L., Doyle, M.,  
 Ferreira, P., Fitzhugh, W., Forrest, C., Funke, R., Gage, D.,  
 Galagan, J., Gardyna, S., Grant, G., Hages, B., Heathford, A., Horton, L.,  
 Howland, J. C., Johnson, R., Jones, C., Kann, L., Karataa, A., Klein, J.,  
 Lenczky, J., Lien, C., Locke, K., MacDonald, P., Marquis, N.,  
 McGarr, P., McGarr, A., McKernan, K., McDonald, J., Meldrum, J.,  
 Morrow, J., Naylor, J., Norman, C. H., O'Connor, T., O'Donnell, P.,  
 Peterson, K., Pollara, V., Riley, R., Roy, A., Santos, R., Severy, P.,  
 Stange-Thomann, N., Stojanovic, N., Subramanian, A., Talamas, J.,  
 Teste, S., Tirrell, A., Vassiliev, H., Vo, A., Wheeler, J., Wu, X.,  
 Wyman, D., Ye, W. J., Zimmer, A. and Zody, M.  
 Direct SubMISSION

TITLE  
 JOURNAL  
 COMMENT  
 Submitted (17-OCT-1999) Whitehead Institute/MIT Center for Genome  
 Research, 320 Charles Street, Cambridge, MA 02141, USA  
 On Apr 21, 2000 this sequence version replaced gi:7008790.  
 All repeats were identified using RepeatMasker:  
 Smit, A.F.A. & Green, P. (1996-1997)  
 http://ftp.genome.washington.edu/RM/RepeatMasker.html

Genome Center  
 Center: Whitehead Institute/ MIT Center for Genome Research  
 Center code: WIRB  
 Web site: http://www-seq.wi.mit.edu  
 Contact: sequence\_submissions@genome.wi.mit.edu  
 Project Information

Center project name: L2592  
 Center clone name: 173\_F\_11  
 Summary Statistics

Sequencing vector: M13; M77815, 100% of reads  
 Chemistry: Dye-terminator: Big Dye, 97% of reads  
 Chemistry: Dye-terminator: Big Dye, 3% of reads  
 Assembly program: Phrap; version 0.960731  
 Consensus quality: 155406 bases at least Q40  
 Consensus quality: 160630 bases at least Q30  
 Consensus quality: 162786 bases at least Q20  
 Insert size: 170000; agarose-fp  
 Insert size: 164301; sum-of-contigs  
 Quality coverage: 4.6 in Q20 bases;  
 Quality coverage: 4.7 in Q20 bases.

NOTE: This is a 'working draft' sequence. It currently  
 consists of 10 contigs. The true order of the pieces  
 is not known and their order in this sequence record is  
 arbitrary. Gaps between the contigs are represented as  
 runs of N, but the exact sizes of the gaps are unknown.  
 This record will be updated with the finished sequence  
 as soon as it is available and the accession number will  
 be preserved.

1. 2517: contig of 2517 bp in length  
 2518 2617: gap of 100 bp  
 2618 5699: contig of 3082 bp in length  
 5700 5799: gap of 100 bp  
 5800 9367: contig of 3568 bp in length  
 9368 9467: gap of 100 bp  
 9468 17246: contig of 7779 bp in length  
 17247 17346: gap of 100 bp  
 17347 28325: contig of 10979 bp in length  
 28326 28425: gap of 100 bp  
 28426 41647: contig of 13222 bp in length  
 41648 41747: gap of 100 bp  
 41748 55693: contig of 13946 bp in length  
 55694 55793: gap of 100 bp  
 55794 79643: contig of 23650 bp in length  
 79644 79743: gap of 100 bp  
 79744 120543: contig of 40800 bp in length  
 120544 120643: gap of 100 bp  
 120644 165201: contig of 44558 bp in length.

## FEATURES

### source

Location/Qualifiers  
 1. 165201  
 /organism="Homo sapiens"  
 /db\_xref="taxon:9606"  
 /chromosome="4"  
 /map="4"  
 /clone="RP11-173F11"  
 /clone\_11b="RP11-173F11 Human Male BAC"  
 1. 2517  
 /note="assembly\_fragment"  
 misc\_feature  
 2618. 5699  
 /note="assembly\_fragment"  
 misc\_feature  
 clone\_end:SP6  
 vector side:right"  
 5800. 9367  
 /note="assembly\_fragment"  
 misc\_feature  
 9468. 17246  
 /note="assembly\_fragment"  
 misc\_feature  
 clone\_end:T7  
 vector side:right"  
 17347. 28325  
 /note="assembly\_fragment"  
 misc\_feature  
 28426. 41647  
 /note="assembly\_fragment"  
 misc\_feature  
 41748. 55693  
 /note="assembly\_fragment"  
 misc\_feature  
 55794. 79643  
 /note="assembly\_fragment"  
 misc\_feature  
 79744. 120543  
 /note="assembly\_fragment"  
 misc\_feature  
 120644. 165201  
 /note="assembly\_fragment"



BASE COUNT 53047 a 31498 c 31586 g 48170 t 900 others  
 ORIGIN

Query Match 2.3%; Score 32; DB 2; Length 165201;  
 Best Local Similarity 100.0%; Pred. No. 5.7e-07;  
 Matches 32; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1358 AACTCTTCCTAATAAAAAAAAAAAAAAAAAA 1389  
 3306 AACTCTTCCTAATAAAAAAAAAAAAAAAAAA 33275

RESULT 15  
 HSA420562 1416 bp mRNA linear PRI 23-NOV-2001  
 LOCUS HSA420562 Homo sapiens mRNA full length insert cDNA clone EUROIMAGE 1525273.  
 DEFINITION  
 AJ420562  
 ACCESSION  
 VERSION AJ420562.1 GI:17066426  
 KEYWORDS  
 SOURCE human.  
 ORGANISM Homo sapiens  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.

REFERENCE 1  
 Auffray C., Ansoerge W., Ballabio A., Estivill X., Gibson K.,  
 Leitch H., Pousetta A. and Lundberg U.  
 TITLE The European IMAGE consortium for integrated Molecular analysis of  
 human gene transcripts  
 JOURNAL Unpublished  
 REFERENCE 2 (bases 1 to 1416)  
 AUTHORS Persson A.  
 TITLE Direct Submission  
 JOURNAL Submitted (02-OCT-2001) Persson A., Center for Molecular  
 Biotechnology, KTH, SCFAB, Institute of Biotechnology, Roslagssvagen  
 30B, 106 91 Stockholm, SWEDEN

COMMENT This clone is available royalty-free through IMAGE Consortium  
 distributors. IMPORTANT: This sequence represents the full insert  
 of this IMAGE cDNA clone. No attempt has been made to verify  
 whether this corresponds to the full-length of the original mRNA  
 from which it was derived.

FEATURES  
 source location/Qualifiers

1..1416  
 /organism="Homo sapiens"  
 /db\_xref="taxon:9606"  
 /chromosome="6"  
 /clone="EUROIMAGE 1525273"  
 /clone\_1lb="Soares\_NFL\_T\_GBC\_S1"  
 polyA\_signal 1357..1362  
 BASE COUNT 484 a 237 c 276 g 419 t  
 ORIGIN

Query Match 2.3%; Score 30; DB 9; Length 1416;  
 Best Local Similarity 100.0%; Pred. No. 7.8e-06;  
 Matches 30; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1361 TCTTCCTAATAAAAAAAAAAAAAAAAAA 1390  
 1375 TCTTCCTAATAAAAAAAAAAAAAAAAAA 1404

Search completed: June 29, 2003, 06:12:36  
 Job time : 2538 secs

GenCore version 5.1.6  
Copyright (c) 1993 - 2003 Compugen Ltd.

OM nucleic - nucleic search, using bw model

Run on: June 29, 2003, 04:14:23 ; Search time 247 Seconds  
(without alignments)  
12737.011 Million cell updates/sec

Title: US-09-989-919-15

Perfect score: 1397  
Sequence: 1 ggtcgtcgcacgtctaccgga.....aaaaaaaaaaagcggtc 1397

Scoring table: OLIGO\_NUC  
Gapop 60.0, Gapext 60.0

Searched: 2185239 seqs, 112599159 residues

Word size: 0

Total number of hits satisfying chosen parameters: 4370478

Minimum DB seq length: 0  
Maximum DB seq length: 200000000

Post-processing: Listing first 1000 summaries

Database: N\_Geneseq\_101002.\*

1: /SID2/gcgdata/geneseq/geneseqn-emb1/NA1980.DAT.\*  
2: /SID2/gcgdata/geneseq/geneseqn-emb1/NA1981.DAT.\*  
3: /SID2/gcgdata/geneseq/geneseqn-emb1/NA1982.DAT.\*  
4: /SID2/gcgdata/geneseq/geneseqn-emb1/NA1983.DAT.\*  
5: /SID2/gcgdata/geneseq/geneseqn-emb1/NA1984.DAT.\*  
6: /SID2/gcgdata/geneseq/geneseqn-emb1/NA1985.DAT.\*  
7: /SID2/gcgdata/geneseq/geneseqn-emb1/NA1986.DAT.\*  
8: /SID2/gcgdata/geneseq/geneseqn-emb1/NA1987.DAT.\*  
9: /SID2/gcgdata/geneseq/geneseqn-emb1/NA1988.DAT.\*  
10: /SID2/gcgdata/geneseq/geneseqn-emb1/NA1989.DAT.\*  
11: /SID2/gcgdata/geneseq/geneseqn-emb1/NA1990.DAT.\*  
12: /SID2/gcgdata/geneseq/geneseqn-emb1/NA1991.DAT.\*  
13: /SID2/gcgdata/geneseq/geneseqn-emb1/NA1992.DAT.\*  
14: /SID2/gcgdata/geneseq/geneseqn-emb1/NA1993.DAT.\*  
15: /SID2/gcgdata/geneseq/geneseqn-emb1/NA1994.DAT.\*  
16: /SID2/gcgdata/geneseq/geneseqn-emb1/NA1995.DAT.\*  
17: /SID2/gcgdata/geneseq/geneseqn-emb1/NA1996.DAT.\*  
18: /SID2/gcgdata/geneseq/geneseqn-emb1/NA1997.DAT.\*  
19: /SID2/gcgdata/geneseq/geneseqn-emb1/NA1998.DAT.\*  
20: /SID2/gcgdata/geneseq/geneseqn-emb1/NA1999.DAT.\*  
21: /SID2/gcgdata/geneseq/geneseqn-emb1/NA2000.DAT.\*  
22: /SID2/gcgdata/geneseq/geneseqn-emb1/NA2001A.DAT.\*  
23: /SID2/gcgdata/geneseq/geneseqn-emb1/NA2001B.DAT.\*  
24: /SID2/gcgdata/geneseq/geneseqn-emb1/NA2002.DAT.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

# SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	1016	72.7	9839	22	AA199052 Human excretory re
2	1016	72.7	9839	22	AA161302 Human kidney relat
3	888	62.7	2461	22	AAH98172 Human EST-derived
4	876	62.7	2461	22	AAH94725 Novel protein kina
5	832	23.0	427	24	ABN94640 Gene #1138 used to
6	813	8.1	150	20	AAH86413 Human single nucle
7	113	8.1	150	20	AAH86414 Human single nucle
8	69	4.9	1837	23	AA578838 DNA encoding novel
9	53	3.8	370	23	AA578836 DNA encoding novel

10	36	2.6	1886	22	AAE44726 Novel protein kina
11	30	2.1	386	24	ABO85200 Arabidopsis thalia
12	30	2.1	421	24	ABL80663 Human ovarian canc
13	30	2.1	777	24	ABN99116 Arabidopsis thalia
14	30	2.1	2457	22	AA545209 Human DNA encoding
15	29	2.1	3339	24	AAH31150 Human carboxylate
16	28	2.0	360	23	ABV49666 Human prostate exp
17	28	2.0	390	23	ABV56432 Human prostate exp
18	28	2.0	458	24	ABT04059 Human ovary specif
19	28	2.0	462	22	AA181254 Human polynucleoti
20	28	2.0	481	23	ABV48841 Human prostate exp
21	28	2.0	642	24	ABQ65026 Arabidopsis thalia
22	28	2.0	905	24	ABT04060 Human ovary specif
23	28	2.0	1292	24	ABL89672 Human polynucleoti
24	28	2.0	1402	21	AA96896 Nucleotide sequenc
25	28	2.0	1485	22	AA527051 cDNA encoding nove
26	28	2.0	1897	21	AAZ43803 Human adult skin c
27	28	2.0	16170	24	ABL33269 Human immune syste
28	28	2.0	37661	22	AAE90334 Human neotokinin B
29	28	2.0	172637	24	ABN83124 Human voltage-acti
30	27	1.9	48	23	ABV46543 Human prostate exp
31	27	1.9	185	23	ABV56476 Human prostate exp
32	27	1.9	207	24	ABL86033 Human ovarian canc
33	27	1.9	258	22	AA110928 Human breast cance
34	27	1.9	297	23	ABV47069 Human prostate exp
35	27	1.9	317	23	ABV49422 Human prostate exp
36	27	1.9	317	23	ABV60613 Human prostate exp
37	27	1.9	325	23	ABV49347 Human prostate exp
38	27	1.9	326	23	ABV47734 Human prostate exp
39	27	1.9	331	23	ABV49436 Human prostate exp
40	27	1.9	344	22	AA188652 Human polynucleoti
41	27	1.9	364	24	ABL85946 Human ovarian canc
42	27	1.9	373	23	ABV48616 Human prostate exp
43	27	1.9	376	23	ABV48965 Human prostate exp
44	27	1.9	386	23	ABV47264 Human prostate exp
45	27	1.9	392	23	ABV46534 Human prostate exp
46	27	1.9	398	23	ABV01976 Human prostate exp
47	27	1.9	404	23	ABV48658 Human prostate exp
48	27	1.9	405	23	ABV45521 Human prostate exp
49	27	1.9	409	22	AA187965 Human polynucleoti
50	27	1.9	411	23	ABV48280 Human prostate exp
51	27	1.9	417	23	ABV58938 Human prostate exp
52	27	1.9	424	23	ABV57393 Human prostate exp
53	27	1.9	428	23	ABV32318 Human prostate exp
54	27	1.9	428	23	ABV32355 Human prostate exp
55	27	1.9	428	23	ABV41197 Human prostate exp
56	27	1.9	428	23	ABV41249 Human prostate exp
57	27	1.9	428	23	ABV44684 Human prostate exp
58	27	1.9	428	23	ABV44709 Human prostate exp
59	27	1.9	428	23	ABV44741 Human prostate exp
60	27	1.9	428	23	ABV44767 Human prostate exp
61	27	1.9	428	23	ABV44799 Human prostate exp
62	27	1.9	433	22	AA191817 Human polynucleoti
63	27	1.9	433	23	ABV47806 Human prostate exp
64	27	1.9	435	22	AA188908 Human polynucleoti
65	27	1.9	435	23	ABV49147 Human prostate exp
66	27	1.9	444	23	ABV32291 Human prostate exp
67	27	1.9	449	23	ABV47740 Human prostate exp
68	27	1.9	450	21	AA655637 Human differentia
69	27	1.9	453	23	ABV49478 Human prostate exp
70	27	1.9	456	23	ABV47610 Human prostate exp
71	27	1.9	456	23	ABV48647 Human prostate exp
72	27	1.9	457	23	ABV48620 Human prostate exp
73	27	1.9	458	23	ABV49161 Human prostate exp
74	27	1.9	458	23	AA583335 DNA encoding novel
75	27	1.9	460	23	ABV46085 Human prostate exp
76	27	1.9	460	23	ABV46381 Human prostate exp
77	27	1.9	460	23	ABV47995 Human prostate exp
78	27	1.9	460	23	ABV48897 Human prostate exp
79	27	1.9	465	23	ABV48253 Human prostate exp
80	27	1.9	467	23	ABV48839 Human prostate exp
81	27	1.9	471	23	ABV48394 Human prostate exp
82	27	1.9	471	23	ABV48730 Human prostate exp

C	83	27	1.9	471	24	ABK54669	Human colon cancer			
	84	27	1.9	474	23	ABV47863	Human prostate exp			
	85	27	1.9	476	23	ABV48137	Human prostate exp			
	86	27	1.9	479	23	ABV49527	Human prostate exp			
	87	27	1.9	486	23	ABV48969	Human prostate exp			
	88	27	1.9	487	23	ABV32296	Human prostate exp			
	89	27	1.9	487	23	ABV41227	Human prostate exp			
	90	27	1.9	487	23	ABV44673	Human prostate exp			
	91	27	1.9	490	23	ABV47641	Human prostate exp			
	92	27	1.9	492	23	ABV47754	Human prostate exp			
	93	27	1.9	494	23	ABV48607	Human prostate exp			
	94	27	1.9	494	23	ABV49669	Human prostate exp			
	95	27	1.9	495	23	ABV48038	Human prostate exp			
	96	27	1.9	495	23	ABV49243	Human prostate exp			
	97	27	1.9	498	23	ABV47956	Human prostate exp			
	98	27	1.9	498	23	ABV48769	Human prostate exp			
	99	27	1.9	502	23	ABV48751	Human prostate exp			
	100	27	1.9	502	23	ABV48780	Human prostate exp			
	101	27	1.9	504	23	ABV49624	Human prostate exp			
	102	27	1.9	504	23	ABV47713	Human prostate exp			
	103	27	1.9	505	23	ABV48972	Human prostate exp			
	104	27	1.9	509	23	ABV47258	Human prostate exp			
	105	27	1.9	509	23	ABV49208	Human prostate exp			
	106	27	1.9	509	23	ABV49636	Human prostate exp			
	107	27	1.9	511	23	ABV48843	Human prostate exp			
	108	27	1.9	511	23	ABV49522	Human prostate exp			
	109	27	1.9	512	23	ABV44744	Human prostate exp			
	110	27	1.9	513	23	ABV48772	Human prostate exp			
	111	27	1.9	516	23	ABV48938	Human prostate exp			
	112	27	1.9	516	23	ABV49485	Human prostate exp			
	113	27	1.9	519	23	ABV47877	Human prostate exp			
	114	27	1.9	524	23	ABV60707	Human prostate exp			
	115	27	1.9	531	23	ABV38011	Human prostate exp			
	116	27	1.9	533	23	ABV49498	Human prostate exp			
	117	27	1.9	534	23	ABV47675	Human prostate exp			
	118	27	1.9	549	23	ABV47750	Human prostate exp			
	119	27	1.9	554	23	ABV46075	Human prostate exp			
	120	27	1.9	556	23	ABV48707	Human prostate exp			
	121	27	1.9	576	23	ABV48943	Human prostate exp			
	122	27	1.9	587	23	ABV48840	Human prostate exp			
	123	27	1.9	588	23	ABV46490	Human prostate exp			
	124	27	1.9	597	23	ABV46464	Human prostate exp			
	125	27	1.9	598	23	ABV48950	Human prostate exp			
	126	27	1.9	608	23	ABV46713	Human prostate exp			
	127	27	1.9	618	22	AAK62213	Human prostate exp			
	128	27	1.9	625	23	ABV46393	Human prostate exp			
	129	27	1.9	637	23	ABV46720	Human prostate exp			
	130	27	1.9	639	23	ABV49431	Human prostate exp			
	131	27	1.9	651	23	ABV46432	Human prostate exp			
	132	27	1.9	653	24	ABO59133	Human prostate exp			
	133	27	1.9	654	23	ABV46386	Human prostate exp			
	134	27	1.9	678	21	AACT9987	Human prostate exp			
	135	27	1.9	684	23	ABV19190	Human prostate exp			
	136	27	1.9	704	24	ABT03074	Human prostate exp			
	137	27	1.9	735	23	ABV49521	Human breast speci			
	138	27	1.9	744	21	AAA47322	Human prostate exp			
	139	27	1.9	773	22	AAI96066	Ftsz1 related sequ			
	140	27	1.9	781	22	AAK71508	Human neuroblastom			
	141	27	1.9	796	22	AAK71507	Human immune/haema			
	142	27	1.9	882	22	AAK45115	Human immune/haema			
	143	27	1.9	938	22	AACT84277	Signal transductio			
	144	27	1.9	1074	21	AACT98014	Human colon cancer			
	145	27	1.9	1142	21	AAZ43924	Human colon cancer			
	146	27	1.9	1291	20	AAK86458	Murine DEDD DNA.			
	147	27	1.9	1301	24	AAK93300	Human prostate a hu			
	148	27	1.9	1338	22	AAK02219	Human prostate spe			
	149	27	1.9	1458	16	AACT9936	Murine Kln17 CDNA.			
	150	27	1.9	1515	21	AACT76840	Human ORF ORF2395			
	151	27	1.9	1516	23	ABV23024	Human prostate exp			
	152	27	1.9	1516	23	ABV25526	Human prostate exp			
	153	27	1.9	1516	23	ABV28860	Human prostate exp			
	154	27	1.9	1674	22	ABV18181	Human phosphatidy			
	155	27	1.9	1817	24	ABK84638	Human CDNA differe			
				156	27	1.9	1817	24	ABL62903	Breast cancer rela
				157	27	1.9	1817	24	ABL63113	Breast cancer rela
				158	27	1.9	2114	20	AAK19958	Rat Tango-76 encod
				159	27	1.9	2290	24	ABK35014	Human CDNA encodin
				160	27	1.9	2417	24	ABK36083	CDNA sequence #474
				161	27	1.9	2491	21	AAK21865	Human breast and o
				162	27	1.9	2752	24	AAK23268	Human Trpbb splice
				163	27	1.9	2803	24	AAK23270	Human Trpbb splice
				164	27	1.9	2908	24	AAK23271	Human Trpbb splice
				165	27	1.9	2913	24	AAK23266	Human Trpba protei
				166	27	1.9	3073	24	AAK23269	Human Trpbb splice
				167	27	1.9	3327	24	ABK91824	DNA representing 1
				168	27	1.9	3780	22	AAK13823	Human encoding huma
				169	27	1.9	4174	22	AAK13821	Human encoding huma
				170	27	1.9	4298	24	AAK34853	Human CDNA for nov
				171	27	1.9	5301	22	AAK97866	Human neuroblastom
				172	27	1.9	5520	24	ABT03038	Human breast speci
				173	27	1.9	5893	24	ABL13858	Human immune syste
				174	27	1.9	6106	24	ABL70545	Chemically treated
				175	27	1.9	6106	24	ABL33076	Human immune syste
				176	27	1.9	6106	24	ABL34522	Human metastasis a
				177	27	1.9	7426	24	ABL46568	DNA encoding G-pro
				178	27	1.9	8167	22	AAK05420	Human reproductive
				179	27	1.9	9539	22	AAK45347	Chemically pretrea
				180	27	1.9	9539	24	ABK28180	DNA transcription
				181	27	1.9	12000	24	ABK34015	Human DNA for stag
				182	27	1.9	12774	22	ABK15302	Human nervous syst
				183	27	1.9	12774	22	AAK72738	Human immune/haema
				184	27	1.9	14562	24	ABQ67048	Human immune/haema
				185	27	1.9	14708	22	AAK45514	Human angiogenesis
				186	27	1.9	14708	24	ABK92325	Chemically pretrea
				187	27	1.9	14708	24	ABL34245	Human immune syste
				188	27	1.9	14708	24	ABK28450	DNA transcription
				189	27	1.9	15000	22	AAK28657	Genomic sequence #
				190	27	1.9	18047	22	AAK80292	Human immune/haema
				191	27	1.9	39353	22	AAK65396	Human immune/haema
				192	27	1.9	39353	22	AAK67232	Human immune/haema
				193	27	1.9	51469	22	AAK69322	Human immune/haema
				194	27	1.9	51469	22	AAK70270	Human immune/haema
				195	27	1.9	51469	22	AAK70813	Human immune/haema
				196	27	1.9	51469	22	AAK78813	Human immune/haema
				197	27	1.9	144460	21	AAZ93815	Olfactory receptor
				198	27	1.9	51	22	AAK130728	Human SNP oligonuc
				199	26	1.9	77	22	AAK56538	Human CDNA for an
				200	26	1.9	107	22	AAH71530	Human cervical can
				201	26	1.9	173	24	ABK45404	CDNA encoding colo
				202	26	1.9	189	24	ABL37790	Human colon tumour
				203	26	1.9	190	23	ABV37682	Human prostate exp
				204	26	1.9	208	22	AAI29051	Colon tumour relat
				205	26	1.9	241	23	ABV61731	Human ovarian can
				206	26	1.9	246	23	ABV61581	Human prostate exp
				207	26	1.9	254	22	AAH82123	Human prostate exp
				208	26	1.9	255	22	AAK24978	Rat differential t
				209	26	1.9	255	22	AAK25293	Human ovarian PCR-
				210	26	1.9	255	22	AAH83623	Human ovarian tumo
				211	26	1.9	267	23	ABV57011	Human prostate exp
				212	26	1.9	277	22	AAH71913	Human cervical can
				213	26	1.9	277	22	AAH14265	Human prostate can
				214	26	1.9	277	23	ABV49364	Human prostate exp
				215	26	1.9	278	23	ABV61306	Human prostate exp
				216	26	1.9	281	24	ABL86992	Human ovarian can
				217	26	1.9	284	23	ABV48902	Human prostate exp
				218	26	1.9	294	22	AAH69989	Human cervical can
				219	26	1.9	299	24	ABL87398	Human ovarian can
				220	26	1.9	300	23	ABV61266	Human prostate exp
				221	26	1.9	302	23	ABV07754	Human prostate exp
				222	26	1.9	312	22	AAH71414	Human cervical can
				223	26	1.9	321	23	ABV48393	Human prostate exp
				224	26	1.9	337	22	ABL10103	Human polynucleoti
				225	26	1.9	337	22	ABL87132	Human ovarian can
				226	26	1.9	352	22	ABL11898	Human breast can
				227	26	1.9	352	22	ABL20788	Human breast can
				228	26	1.9	353	22	AAI87665	Human polynucleoti

C 229	26	1.9	360	22	AA10196	Human breast cancer	302	26	1.9	981	20	AAK87394	Mouse STR1 (Tan
C 230	26	1.9	360	22	AA11976	Human breast cancer	303	26	1.9	1040	24	ABL56860	Mouse X-type secre
C 231	26	1.9	361	23	ABV12942	Human prostate exp	304	26	1.9	1040	24	ABA96630	Mouse X-type secre
C 232	26	1.9	365	23	ABV47637	Human prostate exp	305	26	1.9	1062	22	AAH314142	Human colon cancer
C 233	26	1.9	366	22	AA187465	Human polynucleoti	306	26	1.9	1242	22	AAD02808	HNA040 CDNA/clon
C 234	26	1.9	372	24	ABK64383	Human benign prost	307	26	1.9	1263	21	AAF15596	Human prostate can
C 235	26	1.9	374	22	AA181153	Human polynucleoti	308	26	1.9	1274	24	ABD38087	Corn chlorophylla
C 236	26	1.9	382	23	ABV46589	Human prostate exp	309	26	1.9	1318	24	ABL35068	Marine CDNA isolat
C 237	26	1.9	387	22	AAH71915	Human cervical can	310	26	1.9	1327	20	AAK30162	Human secreted pro
C 238	26	1.9	391	22	AA183087	Human polynucleoti	311	26	1.9	1391	21	AACT7672	Human cancer/assoc
C 239	26	1.9	396	23	ABV49020	Human prostate exp	312	26	1.9	1395	24	ABO51152	Oligonucleotide fo
C 240	26	1.9	397	22	AA180153	Human polynucleoti	313	26	1.9	1395	24	ABO51152	Oligonucleotide fo
C 241	26	1.9	401	22	AA183850	Human polynucleoti	314	26	1.9	1431	22	AAD20262	Rat integrin modu
C 242	26	1.9	402	22	AA187524	Human polynucleoti	315	26	1.9	1509	22	AA158831	Human secreted pro
C 243	26	1.9	402	22	AA185357	Human breast cancer	316	26	1.9	1562	24	ABK69868	Human secreted pro
C 244	26	1.9	403	22	AA185357	Human polynucleoti	317	26	1.9	1562	24	ABK69868	Human secreted pro
C 245	26	1.9	403	22	ABV61005	Human prostate exp	318	26	1.9	1569	21	AAZ98057	Human secreted pro
C 246	26	1.9	405	22	AA119599	Human breast cancer	319	26	1.9	1569	22	AAD11670	Human secreted pro
C 247	26	1.9	405	23	ABV34066	Human prostate exp	320	26	1.9	1569	22	ABK69766	Human secreted pro
C 248	26	1.9	405	23	AAV42931	Human prostate exp	321	26	1.9	1728	23	AAK87014	DNA encoding novel
C 249	26	1.9	413	22	AAK43374	DNA encoding novel	322	26	1.9	1730	21	AAA26383	Human secreted pro
C 250	26	1.9	413	22	AA189438	Human polynucleoti	323	26	1.9	1758	17	AAT32325	Soluble starch syn
C 251	26	1.9	416	23	ABV47951	Human prostate exp	324	26	1.9	1797	21	AAK08375	Rat 4F2 heavy chai
C 252	26	1.9	423	22	AA182334	Human polynucleoti	325	26	1.9	1852	22	AAH18681	Human CDNA sequenc
C 253	26	1.9	427	22	AAD05269	Human secreted pro	326	26	1.9	1960	24	ABN87244	Lolium perenne C4H
C 254	26	1.9	433	22	AA1823127	Human breast cancer	327	26	1.9	2093	17	AAT17864	Hepatitis A virus
C 255	26	1.9	433	22	AA187368	Human polynucleoti	328	26	1.9	2123	24	ABL59578	Rat OST29 gene fra
C 256	26	1.9	437	22	AA118755	Human breast cancer	329	26	1.9	2123	24	ABL59581	Human secreted pro
C 257	26	1.9	440	22	AAH70715	Human cervical can	330	26	1.9	2158	24	ABL50337	Human cancer cell
C 258	26	1.9	441	22	AA119793	Human breast cancer	331	26	1.9	2208	21	AAK59775	Human secreted pro
C 259	26	1.9	442	22	AA183061	Human polynucleoti	332	26	1.9	2288	21	AAK66441	Human secreted pro
C 260	26	1.9	448	22	AA186117	Human polynucleoti	333	26	1.9	2362	22	AA197952	Human neuroblastom
C 261	26	1.9	448	23	ABV49465	Human prostate exp	334	26	1.9	2362	22	AA198105	Human neuroblastom
C 262	26	1.9	448	24	AB187531	Human ovarian can	335	26	1.9	2528	21	AAF16147	Human prostate can
C 263	26	1.9	453	22	AA119861	Human breast cancer	336	26	1.9	2535	22	AAK41164	CDNA encoding nove
C 264	26	1.9	471	24	ABN87880	Human ovary speci	337	26	1.9	2635	22	ABK43877	DNA encoding novel
C 265	26	1.9	473	22	AA190495	Human polynucleoti	338	26	1.9	2660	23	AAH43560	Human DHOR-1 CDNA
C 266	26	1.9	482	23	ABV49211	Human prostate exp	339	26	1.9	2745	23	ABK43576	DNA encoding novel
C 267	26	1.9	504	22	AAK57962	Human prostate exp	340	26	1.9	3001	24	ABK34025	Human DNA for strag
C 268	26	1.9	511	21	AAZ97335	Human immune/haema	341	26	1.9	3149	24	ABK84687	Human neuroblastom
C 269	26	1.9	531	23	ABV46374	Human prostate can	342	26	1.9	3260	22	AAD07596	CDNA encoding nove
C 270	26	1.9	534	23	ABV47036	Human prostate exp	343	26	1.9	3289	21	AAK26281	DNA encoding novel
C 271	26	1.9	540	23	ABV60963	Human prostate exp	344	26	1.9	3304	22	ABAI8014	Human secreted pro
C 272	26	1.9	549	20	AAK87732	Beauveria bassiana	345	26	1.9	3338	21	AAZ94119	Human nervous syst
C 273	26	1.9	578	23	ABV54931	Human prostate exp	346	26	1.9	3408	21	AAZ94096	Haematopoietic ste
C 274	26	1.9	580	22	AAH70713	Human cervical can	347	26	1.9	3536	22	AAD06381	Human ATP binding
C 275	26	1.9	594	24	AA137818	Wheat KCP-like pro	348	26	1.9	3591	20	AAK60932	Murine cell death
C 276	26	1.9	595	22	AAK56873	Human CDNA encodin	349	26	1.9	3712	20	AAK03040	Human IL-1ra BAC c
C 277	26	1.9	600	24	ABO57191	Human colon cancer	350	26	1.9	4116	20	AAK02983	Human IL-1ra BAC c
C 278	26	1.9	619	24	ABO22348	Oligonucleotide fo	351	26	1.9	5640	24	AA169391	Human EDG-8 recept
C 279	26	1.9	619	24	ABO22349	Oligonucleotide fo	352	26	1.9	5647	24	ABL70355	Chemically created
C 280	26	1.9	631	24	ABO25138	Oligonucleotide fo	353	26	1.9	5647	24	ABL33566	Human immune syste
C 281	26	1.9	631	24	ABO25139	Oligonucleotide fo	354	26	1.9	5647	24	AAK561320	Human gene regulat
C 282	26	1.9	632	24	ABO60764	Human colon cancer	355	26	1.9	6086	22	AAK46667	Tumour suppressor
C 283	26	1.9	639	23	ABV19696	Human prostate exp	356	26	1.9	6165	24	ABN97371	Gene #3869 used to
C 284	26	1.9	653	24	ABO57371	Human colon cancer	357	26	1.9	6352	24	ABL32904	Human immune syste
C 285	26	1.9	655	24	AB187011	Human ovarian can	358	26	1.9	6484	24	ABN80138	Human chemically m
C 286	26	1.9	700	22	AAH92102	Human inflammatory	359	26	1.9	6553	22	AAK40055	Genomic sequence #
C 287	26	1.9	711	20	AAK80828	Human secreted pro	360	26	1.9	6553	22	AAK91472	Human digestive sy
C 288	26	1.9	711	21	AAK52476	Human secreted pro	361	26	1.9	6668	24	ABL33697	Human immune syste
C 289	26	1.9	755	24	ABO89771	Human prostate exp	362	26	1.9	6686	24	ABL32461	Human immune syste
C 290	26	1.9	784	24	AAK562079	Porcine musclicar s	363	26	1.9	6854	22	AA137419	Human musculoskele
C 291	26	1.9	785	22	AA195240	Human neuroblastom	364	26	1.9	7921	24	ABL33970	Human immune syste
C 292	26	1.9	785	22	AA198135	Human neuroblastom	365	26	1.9	8912	16	AAT00040	Hepatitis GB virus
C 293	26	1.9	798	22	AA196495	Human neuroblastom	366	26	1.9	8912	21	AAK55280	Hepatitis GB virus
C 294	26	1.9	875	24	ABQ24158	Oligonucleotide fo	367	26	1.9	9139	22	AAK23485	Hepatitis GB virus
C 295	26	1.9	889	22	AA197224	Oligonucleotide fo	368	26	1.9	9143	18	AAT00130	Hepatitis GB virus
C 296	26	1.9	899	20	AA119428	Nicotiana tabacum	369	26	1.9	9143	21	AAK55376	Hepatitis GB virus
C 297	26	1.9	920	22	ABAI8016	Human nervous syst	370	26	1.9	9143	21	AAK55376	Hepatitis GB virus
C 298	26	1.9	927	24	ABT03001	Human breast speci	371	26	1.9	9349	21	AAK55376	Hepatitis GB virus
C 299	26	1.9	944	24	ABK84455	Human CDNA differe	372	26	1.9	9399	22	AAK23484	GBV-B virus genome
C 300	26	1.9	944	24	ABK84455	Human CDNA differe	373	26	1.9	10891	24	ABL32464	Human immune syste
C 301	26	1.9	944	24	ABN98582	Arabidopsis thalia	374	26	1.9	11744	22	AAK83160	Human immune/haema

375	1.9	11853	22	AAK67297	Human immune/haema	448	25	1.8	268	24	ABL86650	Human ovarian canc
376	1.9	11853	22	AAK83162	Human immune/haema	449	25	1.8	269	22	ABL21525	Human breast canc
377	1.9	14753	24	ABA93421	Human alycetamid	450	25	1.8	269	22	ABL22862	Human breast canc
378	1.9	16170	24	ABL33268	Human immune syste	451	25	1.8	276	23	ABL58412	Human prostate exp
379	1.9	20020	22	AAAS29216	Genomic sequence #	452	25	1.8	279	21	AAA44585	Human prostate exp
380	1.9	20020	22	AAAL05535	Human reproductive	453	25	1.8	279	21	ABV49665	Human secreted exp
381	1.9	21423	22	AAAL36470	Human musculoskele	454	25	1.8	281	23	AAAS58627	Human prostate exp
382	1.9	32184	22	ABAA20589	Human nervous syst	455	25	1.8	283	23	AAAS58627	Human prostate exp
383	1.9	32184	22	ABAA20589	Human nervous syst	456	25	1.8	284	23	ABV49472	Human prostate exp
384	1.9	38272	22	AAK84952	Human immune/haema	457	25	1.8	284	23	ABV16736	Human prostate exp
385	1.9	38348	22	AAK84953	Human immune/haema	458	25	1.8	285	23	ABV49372	Human prostate exp
386	1.9	39358	22	AAK65397	Human immune/haema	459	25	1.8	291	20	AAZ13376	Human prostate exp
387	1.9	44848	21	AAAB95533	Nucleotide sequenc	460	25	1.8	291	23	ABV18947	Human gene express
388	1.9	44848	21	AAAB95533	Human corneal N-ac	461	25	1.8	292	20	AAV88635	Est clone HT137.
389	1.9	48436	24	AAAB95533	Human CDNA differe	462	25	1.8	292	23	ABV49135	Human prostate exp
390	1.9	81800	24	ABK84756	Human TBC-1 partia	463	25	1.8	293	23	AAV25018	Human breast canc
391	1.9	99960	21	AAZ50905	Human DNA sequence	464	25	1.8	296	22	AAAL21959	Human breast canc
392	1.9	113515	24	ABL34175	Human glycosyl sul	465	25	1.8	296	24	ABL86554	Human ovarian canc
393	1.9	123219	23	AAH88703	Stomach cancer rel	466	25	1.8	297	24	ABL83791	Human ovarian canc
394	1.9	160552	22	AAAD02697	Thyroid cancer rel	467	25	1.8	297	24	ABL86509	Human ovarian canc
395	1.9	167343	24	ABL64403	pBluescriptSK+ pha	468	25	1.8	297	24	ABL87603	Human ovarian canc
396	1.8	167343	24	ABL67239	Type I RIP gelonin	469	25	1.8	300	22	AAK75362	Human immune/haema
397	1.8	28	21	AAA40362	3'-5' PCR primer X	470	25	1.8	302	23	ABV48079	Human prostate exp
398	1.8	46	16	AAQ75529	Human reproductive	471	25	1.8	303	23	ABV48735	Human prostate exp
399	1.8	46	16	AAQ92339	Melanoma antigen c	472	25	1.8	307	22	ABL37144	Human prostate exp
400	1.8	58	22	AA545526	Linear CDNA librar	473	25	1.8	307	23	ABV18924	Human prostate exp
401	1.8	58	22	AA545526	Human prostate exp	474	25	1.8	307	23	ABV55853	Human prostate exp
402	1.8	62	22	AA545526	Human prostate exp	475	25	1.8	310	23	ABV49033	Human prostate exp
403	1.8	92	21	AAK23718	CDNA encoding nove	476	25	1.8	314	23	ABV39642	Human prostate exp
404	1.8	109	23	ABK41815	Human breast canc	477	25	1.8	314	23	ABV58492	Human prostate exp
405	1.8	135	22	AAAL20071	Human colon tumour	478	25	1.8	314	23	ABV61530	Human prostate exp
406	1.8	138	24	ABL36970	Human breast canc	479	25	1.8	315	23	ABV18604	Human prostate exp
407	1.8	154	22	ABL23561	Human ovarian canc	480	25	1.8	316	22	ABL07941	Human prostate exp
408	1.8	155	24	ABL86320	Human prostate exp	481	25	1.8	319	23	ABV49812	Human prostate exp
409	1.8	161	23	ABV05693	Human prostate exp	482	25	1.8	325	22	AAAS29049	CDNA encoding for
410	1.8	168	24	ABL86276	Human ovarian canc	483	25	1.8	328	22	AAI98806	Human kidney relat
411	1.8	172	22	AAK80646	Human immune/haema	484	25	1.8	328	22	AAI63202	Human testicular a
412	1.8	172	23	ABV57412	Human prostate exp	485	25	1.8	334	22	AAAL05226	Human polynucleoti
413	1.8	173	17	AAAT28143	Sensence-related	486	25	1.8	334	23	ABL98109	Human polynucleoti
414	1.8	173	22	AA560588	Human cancer agent	487	25	1.8	335	22	AAV49238	Human polynucleoti
415	1.8	173	24	ABL86623	Human ovarian canc	488	25	1.8	336	22	AAI80797	Human polynucleoti
416	1.8	176	22	AAU19334	Human prostate exp	489	25	1.8	337	22	AAAS60410	Human cancer agent
417	1.8	181	23	ABV17913	Human ORF324 CDNA	490	25	1.8	337	22	AAI80624	Human polynucleoti
418	1.8	188	24	ABN79377	Human breast agent	491	25	1.8	338	22	AAI80023	Human polynucleoti
419	1.8	189	22	AA560389	Human breast agent	492	25	1.8	338	22	AAV48996	Human prostate exp
420	1.8	189	22	AAI15984	Human immune/haema	493	25	1.8	340	20	AAV89136	Human prostate exp
421	1.8	190	22	AAK66096	Human prostate exp	494	25	1.8	341	24	ABL37571	Human prostate exp
422	1.8	190	23	ABV19361	Breast cancer rela	495	25	1.8	343	22	AAI88728	Human polynucleoti
423	1.8	193	24	ABL62804	Breast cancer rela	496	25	1.8	344	22	AAI14837	Human prostate exp
424	1.8	193	24	ABL62987	Human ovarian canc	497	25	1.8	345	22	AAI23699	Human breast canc
425	1.8	197	24	ABL86648	Human ovarian canc	498	25	1.8	346	22	AAV58481	Human prostate exp
426	1.8	198	24	ABL86620	Human prostate exp	499	25	1.8	348	22	AAI84434	Human polynucleoti
427	1.8	206	22	AAU20128	Human breast agent	500	25	1.8	348	22	ABV03357	Human prostate exp
428	1.8	206	22	AA560654	Human prostate exp	501	25	1.8	349	22	AAI84989	Human polynucleoti
429	1.8	208	23	ABV05696	Human breast canc	502	25	1.8	351	22	AAI86526	Human polynucleoti
430	1.8	209	22	AAAL09007	Human ovarian tumo	503	25	1.8	352	24	ABO85807	Human polynucleoti
431	1.8	209	22	AAH83553	Human ovarian canc	504	25	1.8	354	22	AAAL25687	Human cervicai can
432	1.8	210	24	ABL86394	Human breast canc	505	25	1.8	354	22	AAH71407	Human prostate exp
433	1.8	212	22	AAU19549	Human prostate exp	506	25	1.8	356	23	ABV49201	Human prostate exp
434	1.8	212	23	ABV19390	Sensence-related	507	25	1.8	359	23	ABV34981	Human prostate exp
435	1.8	214	17	AAAT28148	Human cancer agent	508	25	1.8	359	23	ABV48829	Human prostate exp
436	1.8	220	22	AAAL11172	Human prostate exp	509	25	1.8	360	22	AAI86812	Human prostate exp
437	1.8	222	22	AA560391	Human breast agent	510	25	1.8	361	22	AAI88713	Human polynucleoti
438	1.8	228	22	AAI15522	Human prostate exp	511	25	1.8	361	22	AAI88713	Human polynucleoti
439	1.8	231	24	ABK64354	Human prostate exp	512	25	1.8	367	23	ABV19488	Human prostate exp
440	1.8	237	23	ABV61053	Human prostate exp	513	25	1.8	370	22	AA56445	Human prostate exp
441	1.8	241	23	ABV08250	Human prostate exp	514	25	1.8	370	24	ABL87375	Human prostate exp
442	1.8	247	23	ABV19174	Human prostate exp	515	25	1.8	371	22	AAI87277	Human polynucleoti
443	1.8	249	22	AAI11229	Human prostate exp	516	25	1.8	372	22	AAI89007	Human prostate exp
444	1.8	252	23	ABV49870	Rat differential t	517	25	1.8	373	23	ABV14490	Human prostate exp
445	1.8	255	22	AAH82075	Human secreted exp	518	25	1.8	374	23	ABV04389	Human prostate exp
446	1.8	265	21	AAAL2328	Human breast canc	519	25	1.8				
447	1.8	266	22	AAU09162		520	25	1.8				

C 521	25	1.8	377	23	ABV10381	Human prostate exp	C 594	25	1.8	418	23	ABV14865	Human prostate exp
C 522	25	1.8	379	22	AA183173	Human polynucleoti	C 595	25	1.8	422	22	AA187575	Human polynucleoti
C 523	25	1.8	381	22	AAK61735	Human immune/haema	C 596	25	1.8	422	23	ABV43774	Human prostate exp
C 524	25	1.8	381	23	ABV58662	Human prostate exp	C 597	25	1.8	424	22	AA187099	Human polynucleoti
C 525	25	1.8	381	24	ABG85304	Arabidopsis thalia	C 598	25	1.8	425	21	AA187099	Human polynucleoti
C 526	25	1.8	382	22	AA180890	Human polynucleoti	C 599	25	1.8	425	21	AA187099	Human polynucleoti
C 527	25	1.8	382	22	AA182890	Human polynucleoti	C 600	25	1.8	430	21	AA182890	Human polynucleoti
C 528	25	1.8	382	22	AA183090	Human polynucleoti	C 601	25	1.8	430	22	AA182890	Human polynucleoti
C 529	25	1.8	382	22	AA184706	Human polynucleoti	C 602	25	1.8	433	22	AA183978	Human polynucleoti
C 530	25	1.8	382	22	AA185598	Human polynucleoti	C 603	25	1.8	434	23	ABV46404	Human prostate exp
C 531	25	1.8	382	22	AA187606	Human polynucleoti	C 604	25	1.8	435	22	AA182079	Human polynucleoti
C 532	25	1.8	382	23	ABV49116	Human prostate exp	C 605	25	1.8	438	22	AA17497	Human breast cance
C 533	25	1.8	383	22	AA188525	Human polynucleoti	C 606	25	1.8	438	22	AA183772	Human polynucleoti
C 534	25	1.8	384	21	AA189598	Human secreted pro	C 607	25	1.8	439	22	AA183772	Human breast cance
C 535	25	1.8	384	22	AA183079	Human polynucleoti	C 608	25	1.8	440	22	AA191014	Human polynucleoti
C 536	25	1.8	384	22	AA191477	Human polynucleoti	C 609	25	1.8	442	22	AA180488	Human polynucleoti
C 537	25	1.8	385	22	AA187561	Human polynucleoti	C 610	25	1.8	442	23	ABV14233	Human prostate exp
C 538	25	1.8	386	22	AA189019	Human polynucleoti	C 611	25	1.8	443	23	ABV13053	Human prostate exp
C 539	25	1.8	386	22	AA192181	Human polynucleoti	C 612	25	1.8	444	23	ABV56450	Human prostate exp
C 540	25	1.8	388	23	ABV00188	Human prostate exp	C 613	25	1.8	445	24	AB194042	Arabidopsis thalia
C 541	25	1.8	389	22	AA180882	Human breast cance	C 614	25	1.8	446	22	AA183978	Human polynucleoti
C 542	25	1.8	389	23	ABV61827	Human prostate exp	C 615	25	1.8	447	23	ABV13149	Human prostate exp
C 543	25	1.8	390	22	AA184812	Human polynucleoti	C 616	25	1.8	448	22	AA191812	Human polynucleoti
C 544	25	1.8	390	22	AA188511	Human polynucleoti	C 617	25	1.8	448	23	ABV19167	Human prostate exp
C 545	25	1.8	391	22	AA188511	Human polynucleoti	C 618	25	1.8	449	22	AA183772	Human breast cance
C 546	25	1.8	394	22	AA186854	Human polynucleoti	C 619	25	1.8	449	23	ABV42929	Human prostate exp
C 547	25	1.8	395	22	AA181559	Human Mammary Gland	C 620	25	1.8	450	22	AA183053	Human polynucleoti
C 548	25	1.8	396	22	AA188643	Human polynucleoti	C 621	25	1.8	451	23	ABV38949	Human prostate exp
C 549	25	1.8	396	21	AA1844860	Human polynucleoti	C 622	25	1.8	451	23	ABV38952	Human prostate exp
C 550	25	1.8	398	22	AA188703	Human secreted exp	C 623	25	1.8	451	24	ABV058621	Human colon cance
C 551	25	1.8	399	22	AA1883476	Human polynucleoti	C 624	25	1.8	453	22	ABV19043	Human nervous syst
C 552	25	1.8	400	22	AA185595	Human polynucleoti	C 625	25	1.8	454	22	AA183053	Human CDNA for an
C 553	25	1.8	400	22	AA187360	Human polynucleoti	C 626	25	1.8	454	22	AA185101	Human polynucleoti
C 554	25	1.8	400	22	AA187455	Human polynucleoti	C 627	25	1.8	454	23	ABV14338	Human prostate exp
C 555	25	1.8	401	22	AA186707	Human neurotulin g	C 628	25	1.8	454	23	ABV39057	Human prostate exp
C 556	25	1.8	401	22	AA188200	Human neurotulin g	C 629	25	1.8	461	22	AA110372	Human breast cance
C 557	25	1.8	401	22	AA183677	Human polynucleoti	C 630	25	1.8	461	23	ABV39039	Human prostate exp
C 558	25	1.8	403	22	AA182884	Human polynucleoti	C 631	25	1.8	461	23	ABV44778	Human prostate exp
C 559	25	1.8	404	22	AA183869	Human polynucleoti	C 632	25	1.8	463	22	AA183750	Human polynucleoti
C 560	25	1.8	404	23	ABV14862	Human prostate exp	C 633	25	1.8	463	23	ABV18369	Human prostate exp
C 561	25	1.8	404	23	ABV49258	Human prostate exp	C 634	25	1.8	463	22	AA183736	Human polynucleoti
C 562	25	1.8	405	22	AA189652	Human neurotulin g	C 635	25	1.8	464	23	ABV34269	Human prostate exp
C 563	25	1.8	405	22	AA1884779	Human neurotulin g	C 636	25	1.8	464	22	ABV41330	Human prostate exp
C 564	25	1.8	405	22	AA1884779	Human polynucleoti	C 637	25	1.8	466	22	AA182085	Human polynucleoti
C 565	25	1.8	405	23	ABV34180	Human prostate exp	C 638	25	1.8	467	23	ABV45888	Human prostate exp
C 566	25	1.8	405	23	ABV34041	Human prostate exp	C 639	25	1.8	467	24	ABK45850	CDNA encoding colo
C 567	25	1.8	406	19	AAV05122	Nucleotide sequenc	C 640	25	1.8	470	22	AA125852	Human breast cance
C 568	25	1.8	406	22	AA181228	Human polynucleoti	C 641	25	1.8	470	22	AA193681	Human polynucleoti
C 569	25	1.8	406	22	AA185509	Human polynucleoti	C 642	25	1.8	470	23	ABV14765	Human prostate exp
C 570	25	1.8	406	22	AA192822	Human polynucleoti	C 643	25	1.8	470	23	ABV49367	Human prostate exp
C 571	25	1.8	407	23	ABV46364	Human prostate exp	C 644	25	1.8	470	23	ABV61721	Human prostate exp
C 572	25	1.8	407	23	ABV61484	Human prostate exp	C 645	25	1.8	472	23	ABV46233	Human prostate exp
C 573	25	1.8	408	22	AA110289	Human CDNA clone (	C 646	25	1.8	475	23	ABV30533	Human prostate exp
C 574	25	1.8	408	23	ABV49309	Human prostate exp	C 647	25	1.8	475	23	ABV39505	Human prostate exp
C 575	25	1.8	409	22	AA180757	Human polynucleoti	C 648	25	1.8	477	22	AA181157	Human polynucleoti
C 576	25	1.8	410	22	AA184857	Human polynucleoti	C 649	25	1.8	478	21	AA181157	Human breast land o
C 577	25	1.8	410	23	ABV39040	Human prostate exp	C 650	25	1.8	478	23	ABV35326	Human prostate exp
C 578	25	1.8	410	23	ABV44826	Human prostate exp	C 651	25	1.8	478	23	ABV44157	Human prostate exp
C 579	25	1.8	410	23	ABV48387	Human prostate exp	C 652	25	1.8	481	23	ABV44745	Human prostate exp
C 580	25	1.8	411	22	AA184864	Human polynucleoti	C 653	25	1.8	484	23	ABV05169	Human prostate exp
C 581	25	1.8	411	22	AA187361	Human polynucleoti	C 654	25	1.8	484	23	ABV59185	Human prostate exp
C 582	25	1.8	411	22	AA187378	Human polynucleoti	C 655	25	1.8	485	22	AA116432	Human breast cance
C 583	25	1.8	412	22	AA184848	Human polynucleoti	C 656	25	1.8	486	23	ABV35423	Human prostate exp
C 584	25	1.8	412	23	ABV43247	Human prostate exp	C 657	25	1.8	487	23	ABV46451	Human prostate exp
C 585	25	1.8	413	22	AA182940	Human polynucleoti	C 658	25	1.8	489	22	AA119502	Human breast cance
C 586	25	1.8	413	22	AA184095	Human polynucleoti	C 659	25	1.8	493	22	AA119017	Human breast cance
C 587	25	1.8	413	22	AA185119	Human polynucleoti	C 660	25	1.8	495	23	ABV18011	Human prostate exp
C 588	25	1.8	413	23	ABV14974	Human prostate exp	C 661	25	1.8	496	23	ABV48704	Human prostate exp
C 589	25	1.8	414	21	AA181730	Bridge-1 related p	C 662	25	1.8	499	22	AA112544	Human CDNA clone (
C 590	25	1.8	415	22	AA181730	Human polynucleoti	C 663	25	1.8	499	23	ABV46838	Human prostate exp
C 591	25	1.8	415	23	ABV60254	Human prostate exp	C 664	25	1.8	501	22	AA121795	Human breast cance
C 592	25	1.8	416	22	AA191796	Human polynucleoti	C 665	25	1.8	501	22	AA193675	Human polynucleoti
C 593	25	1.8	418	21	AA193675	Human colon cance	C 666	25	1.8	502	22	ABV14084	Human nervous syst

c 667	25	1.8	506	21	AA280662	Human colon cancer
c 668	25	1.8	510	23	ABV48680	Human prostate exp
c 669	25	1.8	510	24	ABN87806	Human ovary specif
c 670	25	1.8	510	24	ABL83785	Human ovarian canc
c 671	25	1.8	515	19	AAV23109	CDNA encoding the
c 672	25	1.8	516	22	AAI25122	Human breast cancer
c 673	25	1.8	516	23	ABV47663	Human prostate exp
c 674	25	1.8	516	23	AAV79165	Human cactexia-ss
c 675	25	1.8	523	18	ABA13577	Human nervous syst
c 676	25	1.8	530	22	AAH10156	Human CDNA clone (
c 677	25	1.8	533	22	AAH34401	Human colon cancer
c 678	25	1.8	533	22	AAH10606	Human CDNA clone (
c 679	25	1.8	535	24	AAH04093	Human ovary specif
c 680	25	1.8	537	23	ABV60414	Human prostate exp
c 681	25	1.8	539	23	ABV57689	Human prostate exp
c 682	25	1.8	540	22	ABF81801	Human secreted pro
c 683	25	1.8	551	24	ABQ58776	Human colon cancer
c 684	25	1.8	552	24	ABQ60515	Human colon cancer
c 685	25	1.8	552	24	ABL65284	Lung cancer relate
c 686	25	1.8	554	23	ABV44736	Human prostate exp
c 687	25	1.8	557	24	ABQ58303	Human colon cancer
c 688	25	1.8	558	23	ABV35701	Human prostate exp
c 689	25	1.8	558	23	ABV44506	Human prostate exp
c 690	25	1.8	560	23	ABV47703	Human prostate exp
c 691	25	1.8	563	22	AAV72800	Secreted protein g
c 692	25	1.8	566	21	AA280382	Human colon cancer
c 693	25	1.8	568	24	ABK54673	Human colon cancer
c 694	25	1.8	571	21	AACT7871	Human cancer assoc
c 695	25	1.8	572	20	AAK63782	Soybean threonine
c 696	25	1.8	573	22	AAK63782	Human immune/haema
c 697	25	1.8	576	21	AAI16432	Human ISF CDNA cto
c 698	25	1.8	580	22	AAQ29218	Oligonucleotide fo
c 699	25	1.8	583	24	ABQ52812	Oligonucleotide fo
c 700	25	1.8	583	24	ABQ52813	Human foecal liver
c 701	25	1.8	587	22	ABQ62563	Probe #8357 for ge
c 702	25	1.8	587	22	ABA29891	Human brain expres
c 703	25	1.8	587	22	AAK10919	Human bone marrow
c 704	25	1.8	587	22	AAK36776	Probe #7560 for ge
c 705	25	1.8	587	22	AAI17627	Probe #11336 used
c 706	25	1.8	587	24	ABQ142550	Human genome-deriv
c 707	25	1.8	587	24	ABQ10780	Human CDNA for nov
c 708	25	1.8	592	21	ABK34545	Human colon cancer
c 709	25	1.8	593	21	AA280444	Human colon cancer
c 710	25	1.8	601	24	ABQ59526	Human ovarian and
c 711	25	1.8	602	22	ABA08020	Human reproductive
c 712	25	1.8	602	22	AAI06681	Human secreted pro
c 713	25	1.8	604	20	AAV79062	Human prostate exp
c 714	25	1.8	606	23	ABV03945	Human prostate exp
c 715	25	1.8	606	23	ABV44739	Human prostate exp
c 716	25	1.8	606	23	ABV48726	Human prostate exp
c 717	25	1.8	608	13	AAQ31796	PRP3 gene (CDNA).
c 718	25	1.8	610	22	AAI19953	Human breast cance
c 719	25	1.8	611	21	AAAI6352	Human colon cancer
c 720	25	1.8	612	24	ABQ60507	Human colon cancer
c 721	25	1.8	612	24	ABL87317	Human ovarian canc
c 722	25	1.8	615	22	AAK88889	Human digestive sy
c 723	25	1.8	615	22	AAI57610	Human colorectal c
c 724	25	1.8	618	24	ABN73814	Bovine embryonic g
c 725	25	1.8	621	22	AA529124	CDNA encoding for
c 726	25	1.8	625	24	ABQ57599	Human colon cancer
c 727	25	1.8	628	23	ABV23066	Human prostate exp
c 728	25	1.8	628	23	ABV23134	Human prostate exp
c 729	25	1.8	628	23	ABV28902	Human prostate exp
c 730	25	1.8	628	23	ABV28975	Human prostate exp
c 731	25	1.8	631	24	ABV61919	Human prostate exp
c 732	25	1.8	632	24	ABK62327	Rat sequence diffe
c 733	25	1.8	635	21	AAQ28042	Human colon cancer
c 734	25	1.8	635	21	AA529885	Human CDNA encodin
c 735	25	1.8	640	22	AAI00884	Human reproductive
c 736	25	1.8	643	22	AA56406	Human CDNA for an
c 737	25	1.8	644	23	ABV13869	Human prostate exp
c 738	25	1.8	645	23	ABV14956	Human prostate exp
c 739	25	1.8	646	20	AAV79022	Human secreted pro
c 740	25	1.8	740	22	AAI01722	Human reproductive
c 741	25	1.8	741	23	ABV05454	Human prostate exp
c 742	25	1.8	742	20	AA217693	Human gene express
c 743	25	1.8	743	25	AAAO2687	Human colon cancer
c 744	25	1.8	744	25	AAV37547	Human secreted pro
c 745	25	1.8	657	20	AAV37547	Human prostate exp
c 746	25	1.8	661	24	ABV14955	Human colon cancer
c 747	25	1.8	661	24	ABO59445	Human colon cancer
c 748	25	1.8	689	22	AAI96694	Human neuroblastom
c 749	25	1.8	690	22	AAI25275	Human breast cance
c 750	25	1.8	695	22	AAI96736	Human neuroblastom
c 751	25	1.8	700	22	AAH92947	Human inflammatory
c 752	25	1.8	701	22	ABAI3166	Human nervous syst
c 753	25	1.8	703	24	AAI42577	Partial sequence o
c 754	25	1.8	705	24	ABI98562	Human neuroblastom
c 755	25	1.8	710	22	AAI97812	Human neuroblastom
c 756	25	1.8	710	22	AAI97860	Human neuroblastom
c 757	25	1.8	710	22	AAI97870	Human neuroblastom
c 758	25	1.8	711	22	AAI97858	Oligonucleotide fo
c 759	25	1.8	713	24	ABQ28962	Oligonucleotide fo
c 760	25	1.8	713	24	ABQ28963	Human ovary specif
c 761	25	1.8	716	24	ABQ57607	Human colon cancer
c 762	25	1.8	716	24	AA596302	Arabidopsis CDNA e
c 763	25	1.8	723	24	AA596330	Human neuroblastom
c 764	25	1.8	723	24	AAI95619	Human neuroblastom
c 765	25	1.8	725	22	AAI95652	Human neuroblastom
c 766	25	1.8	731	22	AAI97146	Human neuroblastom
c 767	25	1.8	732	22	AAH04651	Human CDNA clone (
c 768	25	1.8	736	22	AAI96968	Human neuroblastom
c 769	25	1.8	739	22	AAI97514	Human neuroblastom
c 770	25	1.8	743	21	AAQ59165	Human secreted pro
c 771	25	1.8	745	22	AAH34348	Human colon cancer
c 772	25	1.8	745	22	AAI97227	Human neuroblastom
c 773	25	1.8	747	22	AAI97540	Human neuroblastom
c 774	25	1.8	751	22	AAI97533	Human neuroblastom
c 775	25	1.8	757	22	AAI24366	Human breast cance
c 776	25	1.8	759	22	AAI97528	Human neuroblastom
c 777	25	1.8	766	22	AAI96361	Oligonucleotide fo
c 778	25	1.8	766	22	ABQ18554	Oligonucleotide fo
c 779	25	1.8	769	24	AAI97555	Human neuroblastom
c 780	25	1.8	769	24	AAI97555	Human neuroblastom
c 781	25	1.8	771	22	AAI97078	Oligonucleotide fo
c 782	25	1.8	774	24	ABO52344	Oligonucleotide fo
c 783	25	1.8	774	24	ABO52345	GIPI3/IFI-6-16 ovar
c 784	25	1.8	777	22	AAH83115	Human neuroblastom
c 785	25	1.8	777	22	AAI96237	Flower style-speci
c 786	25	1.8	779	16	AAQ79737	Tomato S-ribonucle
c 787	25	1.8	779	18	AAI64554	Human breast cance
c 788	25	1.8	783	22	AAI95872	Human neuroblastom
c 789	25	1.8	785	24	ABQ27550	Oligonucleotide fo
c 790	25	1.8	788	24	ABO27551	Oligonucleotide fo
c 791	25	1.8	791	22	AAI96087	Human neuroblastom
c 792	25	1.8	792	22	AAI95856	Human neuroblastom
c 793	25	1.8	794	23	AA578134	DNA encoding novel
c 794	25	1.8	795	22	AAH18424	Human CDNA sequenc
c 795	25	1.8	795	24	ABL90186	Human polynucleoti
c 796	25	1.8	797	22	AAI96443	Human neuroblastom
c 797	25	1.8	797	22	AAI94794	Human neuroblastom
c 798	25	1.8	802	22	AAI97441	Human neuroblastom
c 799	25	1.8	805	22	AAK82491	Human immune/haema
c 800	25	1.8	809	22	AAH04384	Human CDNA clone (
c 801	25	1.8	820	22	AAH04384	Human neuroblastom
c 802	25	1.8	828	22	AAI96002	Oligonucleotide fo
c 803	25	1.8	828	24	ABQ14816	Oligonucleotide fo
c 804	25	1.8	828	24	ABQ14817	Oligonucleotide fo
c 805	25	1.8	828	24	ABO46914	Oligonucleotide fo
c 806	25	1.8	829	24	AAH04291	Human CDNA clone (
c 807	25	1.8	829	22	AAZ25123	Rice trypanophan sy
c 808	25	1.8	831	22	AAH64773	Human secreted pro
c 809	25	1.8	831	22	AAH64772	Human secreted pro
c 810	25	1.8	832	22	AAI97461	Human neuroblastom
c 811	25	1.8	833	22	AAI97461	Human neuroblastom
c 812	25	1.8	842	22	AAI94769	Human neuroblastom



C 813	25	1.8	842	22	AAH04491	Human CDNA clone (
C 814	25	1.8	851	23	ABV14623	Human prostate exp
C 815	25	1.8	855	21	AAE21167	Human low adenosis
C 816	25	1.8	855	21	AAA5045	Human adenosis re
C 817	25	1.8	858	22	AAI94029	Human neuroblastom
C 818	25	1.8	865	22	AMS34924	CDNA encoding nove
C 819	25	1.8	867	18	AAH66791	Partial seq. Obtai
C 820	25	1.8	881	22	AAH64790	Human secreted pro
C 821	25	1.8	889	19	AAV34632	Instability prone
C 822	25	1.8	893	10	AAH90956	Nucleotide sequenc
C 823	25	1.8	904	22	AAI97902	Human neuroblastom
C 824	25	1.8	906	21	AAH81046	Human secreted pro
C 825	25	1.8	915	22	AAH07629	Human secreted pro
C 826	25	1.8	920	22	AAI97766	Human neuroblastom
C 827	25	1.8	931	22	AAI97750	Human neuroblastom
C 828	25	1.8	940	18	AAH85879	Malassezia fungus
C 829	25	1.8	941	22	AAI97662	Human neuroblastom
C 830	25	1.8	943	21	AAH59615	Human secreted pro
C 831	25	1.8	946	22	AAH41039	CDNA encoding nove
C 832	25	1.8	952	22	AAI97732	Human neuroblastom
C 833	25	1.8	957	22	AAH33400	Human colon cancer
C 834	25	1.8	958	22	AAI97806	Human neuroblastom
C 835	25	1.8	973	23	AAH87025	DNA encoding novel
C 836	25	1.8	976	22	AAI97748	Human neuroblastom
C 837	25	1.8	1009	24	ABK51969	Wheat CDNA clone w
C 838	25	1.8	1018	21	AAZ34940	Corn anti-silencin
C 839	25	1.8	1024	22	AAH87659	Human immune/haema
C 840	25	1.8	1028	21	AAH97362	Human colorectal c
C 841	25	1.8	1033	20	AAZ77572	Human ovarian tumo
C 842	25	1.8	1034	22	AAH02215	Human reproductive
C 843	25	1.8	1034	22	AAH16213	Human breast or ov
C 844	25	1.8	1073	24	ABH16164	Oligonucleotide fo
C 845	25	1.8	1073	24	ABH16165	Oligonucleotide fo
C 846	25	1.8	1080	24	ABH36744	Oligonucleotide fo
C 847	25	1.8	1080	24	ABH36745	Oligonucleotide fo
C 848	25	1.8	1081	22	AAH08376	Human secreted pro
C 849	25	1.8	1081	22	AAH03919	Human secreted pro
C 850	25	1.8	1081	22	AAH02425	Human secreted pro
C 851	25	1.8	1081	24	ABH19073	Human polynucleoti
C 852	25	1.8	1081	24	ABH01590	Human secreted pro
C 853	25	1.8	1088	18	AAH71217	Type I, p80 IL-1-r
C 854	25	1.8	1101	22	AAH27401	CDNA encoding nove
C 855	25	1.8	1105	20	AAH84989	Human secreted pro
C 856	25	1.8	1110	22	AAH34059	Human colon cancer
C 857	25	1.8	1135	11	AAH03670	Watermelon mosaic
C 858	25	1.8	1144	22	AAH21457	Human nervous syst
C 859	25	1.8	1144	22	AAH84269	Human immune/haema
C 860	25	1.8	1145	22	AAH07658	Human secreted pro
C 861	25	1.8	1153	24	ABH98300	Arabidopsis thalia
C 862	25	1.8	1162	24	ABH054432	Human ovarian anti
C 863	25	1.8	1181	20	AAH62867	Cucumber peroxidase
C 864	25	1.8	1181	20	AAH81616	Cucumber peroxidase
C 865	25	1.8	1189	22	AAH86961	Human immune/haema
C 866	25	1.8	1189	22	AAH86963	Human immune/haema
C 867	25	1.8	1191	22	AAH05511	Human secreted pro
C 868	25	1.8	1198	22	AAH91902	Human secreted pro
C 869	25	1.8	1200	20	AAH74138	Mouse FLAME-2 cDNA
C 870	25	1.8	1210	24	AAH92619	Human class II anti
C 871	25	1.8	1215	22	AAH27964	Arabidopsis thalia
C 872	25	1.8	1218	20	AAH25215	Maize caffeoyl-CoA
C 873	25	1.8	1218	22	AAH05742	Maize caffeoyl-CoA
C 874	25	1.8	1225	20	AAH91408	T. gondii MGIS4-2
C 875	25	1.8	1225	20	AAH91408	T. gondii MGIS4-2
C 876	25	1.8	1225	22	AAH42731	T. gondii CDNA enc
C 877	25	1.8	1225	22	AAH42732	Reverse complement
C 878	25	1.8	1234	21	AAH59835	Human secreted pro
C 879	25	1.8	1239	21	AAH290454	Wnt-4AR and Wnt-5C
C 880	25	1.8	1240	13	AAH024177	Tox2a gene. Pyeno
C 881	25	1.8	1241	19	AAH29479	Nucleic acid encod
C 882	25	1.8	1241	22	AAH07227	Pyemotes tritici p
C 883	25	1.8	1259	24	ABH82131	DNA encoding novel
C 884	25	1.8	1264	24	ABH08208	Human osteoblast d
C 885	25	1.8	1266	24	ABH65866	Lung cancer relate
C 886	25	1.8	1268	24	ABH90630	Human polynucleoti
C 887	25	1.8	1279	24	ABK35913	CDNA sequence #304
C 888	25	1.8	1283	17	AAH30265	Cotton fibre/cell-
C 889	25	1.8	1283	17	AAH13048	Cotton fibre/cell-
C 890	25	1.8	1283	18	AAH70055	Cotton fibre/speci
C 891	25	1.8	1283	18	AAH62624	Cotton fibre/speci
C 892	25	1.8	1287	21	AAH35560	CDNA sequence
C 893	25	1.8	1297	20	AAH24734	Human APP genomic
C 894	25	1.8	1323	24	ABH55101	Human APP sequenc
C 895	25	1.8	1403	22	AAH98071	Human neuroblastom
C 896	25	1.8	1421	20	AAH56000	Human tumour necro
C 897	25	1.8	1436	24	AAH93763	Human signal trans
C 898	25	1.8	1472	22	AAH84196	Human immune/haema
C 899	25	1.8	1480	23	AAH75506	DNA encoding novel
C 900	25	1.8	1480	22	AAH17102	Human CDNA sequenc
C 901	25	1.8	1489	24	ABH14177	Human glycophospha
C 902	25	1.8	1492	23	AAH83131	DNA encoding novel
C 903	25	1.8	1495	21	AAH93115	Human secreted pro
C 904	25	1.8	1498	22	AAH42526	Nucleotide sequenc
C 905	25	1.8	1502	21	AAH75850	Human ORF1405
C 906	25	1.8	1516	22	AAH41436	CDNA encoding nove
C 907	25	1.8	1526	24	AAH01977	Soybean biotin syn
C 908	25	1.8	1529	19	AAH07908	Human cytochrome-r
C 909	25	1.8	1536	21	AAH75132	CDNA encoding a hu
C 910	25	1.8	1564	21	AAH26346	Human secreted pro
C 911	25	1.8	1571	20	AAH06546	Lazy 1a1 gene of m
C 912	25	1.8	1573	24	ABH86082	Human caudal factor
C 913	25	1.8	1576	22	AAH24154	Human secreted pro
C 914	25	1.8	1580	21	AAH61288	Human secreted pro
C 915	25	1.8	1592	22	AAH26905	Human HS-glycoprot
C 916	25	1.8	1593	24	ABH55089	Human CDNA sequenc
C 917	25	1.8	1594	24	ABH55089	Human ovarian anti
C 918	25	1.8	1613	22	AAH03998	Human protein tyro
C 919	25	1.8	1628	22	AAH86029	Human immune/haema
C 920	25	1.8	1631	23	ABH21694	Human prostate exp
C 921	25	1.8	1631	23	ABH27515	Human prostate exp
C 922	25	1.8	1647	20	AAH24036	Human endometrium
C 923	25	1.8	1651	21	AAH257845	Protein regulating
C 924	25	1.8	1656	22	AAH21363	Human CDNA sequenc
C 925	25	1.8	1656	24	ABH85708	DNA encoding Cadu
C 926	25	1.8	1721	22	AAH45012	CDNA encoding nove
C 927	25	1.8	1731	20	AAH98008	CDNA encoding huma
C 928	25	1.8	1736	24	ABH51381	Human prostate exp
C 929	25	1.8	1745	23	ABH25601	Human prostate exp
C 930	25	1.8	1752	23	ABH22170	Human prostate exp
C 931	25	1.8	1752	23	ABH28009	Human prostate exp
C 932	25	1.8	1772	22	AAH16504	Human CDNA sequenc
C 933	25	1.8	1804	24	ABH55667	Phosphoribose glyco
C 934	25	1.8	1829	19	AAH26298	Phosphate stativat
C 935	25	1.8	1838	22	AAH68572	Human immune/haema
C 936	25	1.8	1838	24	ABH39589	Human cancer (suppr
C 937	25	1.8	1847	22	AAH18724	Human CDNA sequenc
C 938	25	1.8	1857	20	AAH21532	Progression elevat
C 939	25	1.8	1868	24	ABH01581	Human secreted pro
C 940	25	1.8	1869	22	AAH15827	Human CDNA sequenc
C 941	25	1.8	1875	18	AAH48669	Human EDG-2 recept
C 942	25	1.8	1893	22	AAH05084	Human secreted pro
C 943	25	1.8	1899	21	AAH98962	Human pancreatic c
C 944	25	1.8	1935	23	ABH30101	Human prostate exp
C 945	25	1.8	1942	21	AAH48939	Soybean diacylglyc
C 946	25	1.8	1948	21	AAH81726	Human secreted pro
C 947	25	1.8	1958	21	AAH81734	Human secreted pro
C 948	25	1.8	1958	22	AAH33172	Human colon cancer
C 949	25	1.8	1959	22	AAH86468	Human cell death p
C 950	25	1.8	1960	19	AAH34257	Med-related human
C 951	25	1.8	1965	22	AAH84281	Human secreted pro
C 952	25	1.8	1981	18	AAH72748	Human prostate can
C 953	25	1.8	1992	22	AAH24555	Zinc finger protei
C 954	25	1.8	2004	22	AAH24555	Oligonucleotide fo
C 955	25	1.8	2005	24	ABH36550	Oligonucleotide fo
C 956	25	1.8	2005	24	ABH36551	Oligonucleotide fo
C 957	25	1.8	2005	24	ABH39768	Oligonucleotide fo
C 958	25	1.8	2005	24	ABH39769	Oligonucleotide fo



959	25	1.8	2007	24	ABO54845	Human ovarian anti
960	25	1.8	2010	24	AAS94721	Rat secreted facto
961	25	1.8	2010	22	AAH33324	Human colon cancer
962	25	1.8	2040	22	AAF25729	C. purpureus delta
963	25	1.8	2053	22	AA541397	CDNA encoding nove
964	25	1.8	2054	21	AAZ43779	Human fecal brain
965	25	1.8	2074	20	AAK04354	Human secreted pro
966	25	1.8	2085	22	AAH35013	Human colon cancer
967	25	1.8	2120	22	AA503916	Human secreted pro
968	25	1.8	2120	22	AA503916	Human polynucleoti
969	25	1.8	2122	24	ABL89840	Human lung-specific
970	25	1.8	2124	22	AAH34555	Human colon cancer
971	25	1.8	2144	22	AAI97916	Human neuroblastom
972	25	1.8	2144	22	AAI98072	Human neuroblastom
973	25	1.8	2152	22	ABA21249	Human nervous syst
974	25	1.8	2173	24	ABK34604	Human CDNA for nov
975	25	1.8	2175	21	AAZ93714	F-box protein FMD1
976	25	1.8	2202	22	AAH17401	Human CDNA sequenc
977	25	1.8	2216	23	ABV23716	Human prostate exp
978	25	1.8	2216	23	ABV23716	Human prostate exp
979	25	1.8	2221	22	ABA19249	Human nervous syst
980	25	1.8	2223	22	AA531296	Human CDNA encodin
981	25	1.8	2223	24	ABO66620	Human polynucleoti
982	25	1.8	2228	22	ABA19250	Human nervous syst
983	25	1.8	2229	21	AAI77766	Human cancer assoc
984	25	1.8	2231	24	ABO54629	Human ovarian anti
985	25	1.8	2237	22	AAI97907	Human neuroblastom
986	25	1.8	2238	22	AAI98059	Human neuroblastom
987	25	1.8	2242	23	ABK43485	DNA encoding novel
988	25	1.8	2242	24	ABL90223	Human polynucleoti
989	25	1.8	2251	20	AAH87414	Hepelocellular car
990	25	1.8	2262	22	AAI98095	Human neuroblastom
991	25	1.8	2263	20	AAH44435	Human brain CNG-1
992	25	1.8	2273	21	AACT7863	Human cancer assoc
993	25	1.8	2285	14	AAQ40574	Gp85-97 clone 18 c
994	25	1.8	2290	24	ABK35014	Human CDNA encodin
995	25	1.8	2295	22	AAI05137	Human reproductive
996	25	1.8	2322	21	AACT9979	Human secreted pro
997	25	1.8	2332	20	AAH33227	Wheat viviparus 1
998	25	1.8	2334	22	AAH14844	Human CDNA sequenc
999	25	1.8	2340	22	AAH33981	Human colon cancer
1000	25	1.8	2342	21	AACT9083	Human pancreatic c

## ALIGNMENTS

## RESULT 1

AAI99052 standard; DNA; 9839 BP.

AAI99052;

07-JAN-2002 (first entry)

Human excretory related polynucleotide SEQ ID NO 816.

XX Human; noctropic; neuroprotective; cytostatic; dermatological; vituicide;  
XX immunosuppressive; antiinflammatory; anti-HIV; antibacterial; vulnery;  
XX antiparkinsonian; antitickling; antianemic; antitachycardic; cancer;  
XX antineuritic; hepatotropic; cerebroprotective; antiinflammatory;  
XX antiallergic; antidiabetic; antitumor; anticonvulsant; antifungal;  
XX antiparasitic; cardiac; immune disorder; cardiovascular disorder;  
XX neurological disease; infection; nephrotropic; gene therapy; vaccine;  
XX excretory system; ds.  
XX Homo sapiens.  
XX WO200155313-A2.  
XX 02-AUG-2001.  
XX 17-JAN-2001; 2001WO-US01323.  
PF

XX 31-JAN-2000; 2000US-0179065.  
PR 04-FEB-2000; 2000US-0180628.  
PR 24-FEB-2000; 2000US-0184664.  
PR 02-MAR-2000; 2000US-0186350.  
PR 16-MAR-2000; 2000US-0189874.  
PR 17-MAR-2000; 2000US-0190076.  
PR 18-APR-2000; 2000US-0198123.  
PR 19-MAY-2000; 2000US-0205515.  
PR 07-JUN-2000; 2000US-0209486.  
PR 28-JUN-2000; 2000US-0214886.  
PR 30-JUN-2000; 2000US-0215133.  
PR 07-JUL-2000; 2000US-0216647.  
PR 07-JUL-2000; 2000US-0216880.  
PR 11-JUL-2000; 2000US-0217487.  
PR 11-JUL-2000; 2000US-0217486.  
PR 14-JUL-2000; 2000US-0218290.  
PR 26-JUL-2000; 2000US-0220963.  
PR 26-JUL-2000; 2000US-0220964.  
PR 14-AUG-2000; 2000US-0224518.  
PR 14-AUG-2000; 2000US-0224519.  
PR 14-AUG-2000; 2000US-0225213.  
PR 14-AUG-2000; 2000US-0225214.  
PR 14-AUG-2000; 2000US-0225266.  
PR 14-AUG-2000; 2000US-0225267.  
PR 14-AUG-2000; 2000US-0225270.  
PR 14-AUG-2000; 2000US-0225447.  
PR 14-AUG-2000; 2000US-0225757.  
PR 14-AUG-2000; 2000US-0225758.  
PR 14-AUG-2000; 2000US-0225759.  
PR 14-AUG-2000; 2000US-0225759.  
PR 18-AUG-2000; 2000US-0226279.  
PR 22-AUG-2000; 2000US-0226581.  
PR 22-AUG-2000; 2000US-0226586.  
PR 22-AUG-2000; 2000US-0227182.  
PR 23-AUG-2000; 2000US-0227009.  
PR 30-AUG-2000; 2000US-0228924.  
PR 01-SEP-2000; 2000US-0229287.  
PR 01-SEP-2000; 2000US-0229343.  
PR 01-SEP-2000; 2000US-0229344.  
PR 01-SEP-2000; 2000US-0229345.  
PR 05-SEP-2000; 2000US-0229509.  
PR 05-SEP-2000; 2000US-0229513.  
PR 06-SEP-2000; 2000US-0230437.  
PR 06-SEP-2000; 2000US-0230438.  
PR 06-SEP-2000; 2000US-0231242.  
PR 08-SEP-2000; 2000US-0231243.  
PR 08-SEP-2000; 2000US-0231244.  
PR 08-SEP-2000; 2000US-0231413.  
PR 08-SEP-2000; 2000US-0231414.  
PR 08-SEP-2000; 2000US-0232080.  
PR 08-SEP-2000; 2000US-0232081.  
PR 12-SEP-2000; 2000US-0231968.  
PR 14-SEP-2000; 2000US-0232397.  
PR 14-SEP-2000; 2000US-0232398.  
PR 14-SEP-2000; 2000US-0232399.  
PR 14-SEP-2000; 2000US-0232400.  
PR 14-SEP-2000; 2000US-0232401.  
PR 14-SEP-2000; 2000US-0233063.  
PR 14-SEP-2000; 2000US-0233064.  
PR 14-SEP-2000; 2000US-0233065.  
PR 21-SEP-2000; 2000US-0234223.  
PR 21-SEP-2000; 2000US-0234274.  
PR 25-SEP-2000; 2000US-0234997.  
PR 25-SEP-2000; 2000US-0234998.  
PR 26-SEP-2000; 2000US-0235484.  
PR 27-SEP-2000; 2000US-0235634.  
PR 27-SEP-2000; 2000US-0235636.  
PR 29-SEP-2000; 2000US-0236327.  
PR 29-SEP-2000; 2000US-0236367.  
PR 29-SEP-2000; 2000US-0236368.  
PR 29-SEP-2000; 2000US-0236369.  
PR 29-SEP-2000; 2000US-0236370.



```
QY 722 CTAGCTATGGAGTCTGCTGCTCTCTAGTCCAGGATCATGGGGGATGATGATGCTCTCCAA 781
DB 9192 CTAGCTATGGAGTCTGCTGCTCTCTAGTCCAGGATCATGGGGGATGATGATGCTCTCCAA 9251
QY 782 CCCCTGAGGCTGTAGCAAGCTCAGGCTAGTCTCCCACTGGGGGCTGTCCTCCCTG 841
DB 9252 CCCCTGAGGCTGTAGCAAGCTCAGGCTAGTCTCCCACTGGGGGCTGTCCTCCCTG 9310
QY 842 GAGAGGTTCCGTTGGGGGAGCCCATCACTGTGTTCATATGAGAAATGAGTAAAGCC 901
DB 9311 GAGAGGTTCCGTTGGGGGAGCCCATCACTGTGTTCATATGAGAAATGAGTAAAGCC 9370
QY 902 CTTGCTGCTGCTGCTGCAATGCCACAGAGCGGCTGGGGGCTGCTGGGAGCAATCCAT 961
DB 9371 CTTGCTGCTGCTGCTGCAATGCCACAGAGCGGCTGGGGGCTGCTGGGAGCAATCCAT 9430
QY 962 CGTGGAGTGTCTCTCAGCTTAGGTGAGCAGAGAGCTGGGGGAGTGTCCAGGAT 1021
DB 9431 CGTGGAGTGTCTCTCAGCTTAGGTGAGCAGAGAGCTGGGGGAGTGTCCAGGAT 9490
QY 1022 GTGGGATATCTGATCCTGGGGAGGCTATCTGTGACCTCCCGACAGGGGACACTCCAG 1081
DB 9491 GTGGGATATCTGATCCTGGGGAGGCTATCTGTGACCTCCCGACAGGGGACACTCCAG 9550
QY 1082 CCAGCCCAAGGGGCTCAGGGGCAAGGTGCAACCTCAGCATGAGCCAAACTGGGGCTCAG 1141
DB 9551 CCAGCCCAAGGGGCTCAGGGGCAAGGTGCAACCTCAGCATGAGCCAAACTGGGGCTCAG 9610
QY 1142 GAGAGGTTGTGTGAGGCAAGGACCTGGGGGCGGGGCTGGGGGCGCTTCTGCTC 1201
DB 9611 GAGAGGTTGTGTGAGGCAAGGACCTGGGGGCGGGGCTGGGGGCGCTTCTGCTC 9670
QY 1202 ATTTGCTTTCAATAAAGCCTCAAGAGCCCAAAACAGGCTTCCCTCTCTCAGATT 1261
DB 9671 ATTTGCTTTCAATAAAGCCTCAAGAGCCCAAAACAGGCTTCCCTCTCTCAGATT 9730
QY 1262 TGAATATCCAGAACTTTTGTACTCTTGTGTGTTAATTTGTATTTTGTAAATA 1321
DB 9731 TGAATATCCAGAACTTTTGTACTCTTGTGTGTTAATTTGTATTTTGTAAATA 9790
QY 1322 AAATAAATATGTAATAAATGATGTTTCACAGCAAACTTCCCTAA 1370
DB 9791 AAATAAATATGTAATAAATGATGTTTCACAGCAAACTTCCCTAA 9839

RESULT 2
AAI63402
ID AAI63402 standard; DNA; 9839 BP.
XX
AC AAI63402;
AC
DT 22-OCT-2001 (first entry)
XX
DE Human kidney related polynucleotide SEQ ID NO 717.
XX
KW Human; kidney antigen; immunosuppressive; antiarthritic; antineumatic;
KW antiproliferative; cytostatic; cardiant; vasotropic; cerebrioprotective;
KW neurotropic; neuroprotective; antibacterial; vitinoid; fungicide;
KW ophthalmological; antiallergic; hepatotropic; antidiabetic;
KW antiinflammatory; antitumor; anticonvulsant; antiparasitic;
KW gene therapy; cancer; immune disorder; cardiovascular disorder;
KW neurological disease; infection; ds.
XX
XX Homo sapiens.
XX OS
XX PN WO200155323-A2.
XX
XX 02-AUG-2001.
XX
XX 17-JAN-2001; 2001WO-US01343.
XX
XX 31-JAN-2000; 2000US-0179065.
PR
```

```
PR 04-FEB-2000; 2000US-0180628.
PR 24-FEB-2000; 2000US-0184664.
PR 02-MAR-2000; 2000US-0186350.
PR 16-MAR-2000; 2000US-0188874.
PR 17-MAR-2000; 2000US-0190076.
PR 18-APR-2000; 2000US-0198123.
PR 19-MAY-2000; 2000US-0205515.
PR 07-JUN-2000; 2000US-0209467.
PR 26-JUN-2000; 2000US-0214886.
PR 30-JUN-2000; 2000US-0215135.
PR 07-JUL-2000; 2000US-0216647.
PR 07-JUL-2000; 2000US-0216880.
PR 11-JUL-2000; 2000US-0217487.
PR 11-JUL-2000; 2000US-0217496.
PR 14-JUL-2000; 2000US-0218290.
PR 26-JUL-2000; 2000US-0220963.
PR 26-JUL-2000; 2000US-0220964.
PR 14-AUG-2000; 2000US-0224518.
PR 14-AUG-2000; 2000US-0224519.
PR 14-AUG-2000; 2000US-0225213.
PR 14-AUG-2000; 2000US-0225214.
PR 14-AUG-2000; 2000US-0225266.
PR 14-AUG-2000; 2000US-0225267.
PR 14-AUG-2000; 2000US-0225268.
PR 14-AUG-2000; 2000US-0225270.
PR 14-AUG-2000; 2000US-0225447.
PR 14-AUG-2000; 2000US-0225757.
PR 14-AUG-2000; 2000US-0225758.
PR 14-AUG-2000; 2000US-0225759.
PR 18-AUG-2000; 2000US-0226279.
PR 22-AUG-2000; 2000US-0226681.
PR 22-AUG-2000; 2000US-0226688.
PR 22-AUG-2000; 2000US-0227182.
PR 23-AUG-2000; 2000US-0227009.
PR 30-AUG-2000; 2000US-0228924.
PR 01-SEP-2000; 2000US-0229287.
PR 01-SEP-2000; 2000US-0229343.
PR 01-SEP-2000; 2000US-0229344.
PR 01-SEP-2000; 2000US-0229345.
PR 05-SEP-2000; 2000US-0229509.
PR 05-SEP-2000; 2000US-0229513.
PR 06-SEP-2000; 2000US-0230437.
PR 06-SEP-2000; 2000US-0230438.
PR 08-SEP-2000; 2000US-0231242.
PR 08-SEP-2000; 2000US-0231243.
PR 08-SEP-2000; 2000US-0231244.
PR 08-SEP-2000; 2000US-0231413.
PR 08-SEP-2000; 2000US-0231414.
PR 08-SEP-2000; 2000US-0232080.
PR 08-SEP-2000; 2000US-0232081.
PR 12-SEP-2000; 2000US-0232968.
PR 14-SEP-2000; 2000US-0232997.
PR 14-SEP-2000; 2000US-0232998.
PR 14-SEP-2000; 2000US-0232999.
PR 14-SEP-2000; 2000US-0232999.
PR 14-SEP-2000; 2000US-0232400.
PR 14-SEP-2000; 2000US-0232401.
PR 14-SEP-2000; 2000US-0233063.
PR 14-SEP-2000; 2000US-0233064.
PR 14-SEP-2000; 2000US-0233065.
PR 21-SEP-2000; 2000US-0234223.
PR 21-SEP-2000; 2000US-0234274.
PR 25-SEP-2000; 2000US-0234997.
PR 25-SEP-2000; 2000US-0234998.
PR 25-SEP-2000; 2000US-0234998.
PR 26-SEP-2000; 2000US-0235484.
PR 27-SEP-2000; 2000US-0235634.
PR 27-SEP-2000; 2000US-0235636.
PR 29-SEP-2000; 2000US-0236327.
PR 29-SEP-2000; 2000US-0236367.
PR 29-SEP-2000; 2000US-0236368.
PR 29-SEP-2000; 2000US-0236369.
PR 29-SEP-2000; 2000US-0236370.
PR 02-OCT-2000; 2000US-0236802.
PR 02-OCT-2000; 2000US-0237037.
```

PR 02-OCT-2000; 2000US-0237038.  
PR 02-OCT-2000; 2000US-0237039.  
PR 02-OCT-2000; 2000US-0237040.  
PR 13-OCT-2000; 2000US-0239935.  
PR 13-OCT-2000; 2000US-0239937.  
PR 20-OCT-2000; 2000US-0240960.  
PR 20-OCT-2000; 2000US-0241221.  
PR 20-OCT-2000; 2000US-0241785.  
PR 20-OCT-2000; 2000US-0241786.  
PR 20-OCT-2000; 2000US-0241787.  
PR 20-OCT-2000; 2000US-0241808.  
PR 20-OCT-2000; 2000US-0241809.  
PR 20-OCT-2000; 2000US-0241826.  
PR 01-NOV-2000; 2000US-0246174.  
PR 08-NOV-2000; 2000US-0246474.  
PR 08-NOV-2000; 2000US-0246475.  
PR 08-NOV-2000; 2000US-0246476.  
PR 08-NOV-2000; 2000US-0246477.  
PR 08-NOV-2000; 2000US-0246478.  
PR 08-NOV-2000; 2000US-0246523.  
PR 08-NOV-2000; 2000US-0246524.  
PR 08-NOV-2000; 2000US-0246525.  
PR 08-NOV-2000; 2000US-0246526.  
PR 08-NOV-2000; 2000US-0246527.  
PR 08-NOV-2000; 2000US-0246528.  
PR 08-NOV-2000; 2000US-0246532.  
PR 08-NOV-2000; 2000US-0246609.  
PR 08-NOV-2000; 2000US-0246610.  
PR 08-NOV-2000; 2000US-0246611.  
PR 08-NOV-2000; 2000US-0246613.  
PR 17-NOV-2000; 2000US-0249207.  
PR 17-NOV-2000; 2000US-0249208.  
PR 17-NOV-2000; 2000US-0249209.  
PR 17-NOV-2000; 2000US-0249210.  
PR 17-NOV-2000; 2000US-0249211.  
PR 17-NOV-2000; 2000US-0249212.  
PR 17-NOV-2000; 2000US-0249213.  
PR 17-NOV-2000; 2000US-0249214.  
PR 17-NOV-2000; 2000US-0249215.  
PR 17-NOV-2000; 2000US-0249216.  
PR 17-NOV-2000; 2000US-0249217.  
PR 17-NOV-2000; 2000US-0249218.  
PR 17-NOV-2000; 2000US-0249244.  
PR 17-NOV-2000; 2000US-0249245.  
PR 17-NOV-2000; 2000US-0249264.  
PR 17-NOV-2000; 2000US-0249265.  
PR 17-NOV-2000; 2000US-0249297.  
PR 17-NOV-2000; 2000US-0249299.  
PR 17-NOV-2000; 2000US-0249300.  
PR 01-DEC-2000; 2000US-0250160.  
PR 01-DEC-2000; 2000US-0250391.  
PR 05-DEC-2000; 2000US-0251030.  
PR 05-DEC-2000; 2000US-0251988.  
PR 05-DEC-2000; 2000US-0256719.  
PR 06-DEC-2000; 2000US-0251479.  
PR 08-DEC-2000; 2000US-0251856.  
PR 08-DEC-2000; 2000US-0251868.  
PR 08-DEC-2000; 2000US-0251869.  
PR 08-DEC-2000; 2000US-0251889.  
PR 08-DEC-2000; 2000US-0251990.  
PR 11-DEC-2000; 2000US-0254097.  
PR 05-JAN-2001; 2001US-0259678.  
XX  
XX (HUMA-) HUMAN GENOME SCI INC.  
XX  
XX Rosen CA, Barash SC, Ruben SM;  
XX  
XX MPI; 2001-488784/53.  
XX  
XX New isolated nucleic acids and polypeptides, useful for diagnosing,  
PT treating and/or preventing human diseases and disorders -  
XX  
XX Disclosure; SEQ ID NO 717; 564pp + Sequence Listing; English.

XX  
CC The invention relates to novel kidney related polynucleotides  
CC (AA162971-AA163793) and the encoded polypeptides (AAM2417-AAM4261)  
CC collectively known as kidney antigens and the use of such kidney antigens  
CC for detecting disorders of the kidney, especially kidney cancer and  
CC kidney cancer metastases. The polynucleotides and proteins are also  
CC useful for preventing, treating or ameliorating medical conditions  
CC e.g. by protein or gene therapy. The genes are isolated from a range  
CC of human tissues disclosed in the specification. The nucleic acids,  
CC proteins, antibodies and (antagonists are useful in the diagnosis,  
CC treatment and prevention of: (a) cancer, e.g. breast and ovarian cancer,  
CC and other cancers of the adrenal gland, bone, bone marrow, breast,  
CC gastrointestinal tract, liver, lung, or urogenital; (b) immune disorders  
CC e.g. Addison's disease, allergies, autoimmune hemolytic anaemia,  
CC autoimmune thyroiditis, diabetes mellitus, Crohn's disease, multiple  
CC sclerosis, rheumatoid arthritis and ulcerative colitis;  
CC (c) cardiovascular disorders such as myocardial ischaemias; (d) wound  
CC healing; (e) neurological diseases e.g. cerebral anoxia and epilepsy;  
CC and (f) infectious diseases such as viral, bacterial, fungal and  
CC parasitic infections.  
CC Note: The sequence data for this patent did not form part of the  
CC printed specification, but was obtained in electronic format directly  
CC from WIPO at ftp.wipo.int/pub/published\_pct\_sequences.  
XX  
XX Sequence 9839 BP; 2292 A; 2702 C; 2654 G; 2191 T; 0 other;  
SQ  
Query Match 72.7%; Score 1016; DB 22; Length 9839;  
Best Local Similarity 99.8%; Pred. No. 0;  
Matches 1306; Conservative 0; Mismatches 1; Indels 2; Gaps 2;  
QY 62 GAGTACCACTGATATCCACAGACACACATCCCCAGAGAGACTACCGTCTGGCCATCC 121  
DB 8533 GAGTACCACTGATATCCACAGACACACATCCCCAGAGAGACTACCGTCTGGCCATCC 8592  
QY 122 TACCAACACGAGAGCTGCTCTTTCAGTGTTCACCTGAGCTGTGATGTCTGT 181  
DB 8593 TACCAACACGAGAGCTGCTCTTTCAGTGTTCACCTGAGCTGTGATGTCTGT 8652  
QY 182 GAGAGCCATGCCAGATGCGGGCTTTGTGTGACCAACACACCTGAGAGCTGTG 241  
DB 8653 GAGAGCCATGCCAGATGCGGGCTTTGTGTGACCAACACACCTGAGAGCTGTG 8712  
QY 242 CCAAGTGGAGAGACCTTTCAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 301  
DB 8713 CCAAGTGGAGAGACCTTTCAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 8772  
QY 302 GATCCCTCATTCGAAAGTCAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 361  
DB 8773 GATCCCTCATTCGAAAGTCAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 8832  
QY 362 CAAGGAGTCAATTAAG 421  
DB 8833 CAAGGAGTCAATTAAG 8892  
QY 422 TACCTCTCACTCTGCCCCCAGAGTGGAGAGTGTCTTTTCAAGAGCTGATGAGCA 481  
DB 8893 TACCTCTCACTCTGCCCCCAGAGTGGAGAGTGTCTTTTCAAGAGCTGATGAGCA 8952  
QY 482 AGTGTCTCTGATATCCACAGACCAATATGTGAATGCTGTGCTGATCTATCTGAG 541  
DB 8953 AGTGTCTCTGATATCCACAGACCAATATGTGAATGCTGTGCTGATCTATCTGAG 9011  
QY 542 GCTCGGCTGACAGAGTGAATATCTCAGAGAGTGGCTTGTGAGAGAGAGTGA 601  
DB 9012 GCTCGGCTGACAGAGTGAATATCTCAGAGAGTGGCTTGTGAGAGAGAGTGA 9071  
QY 602 CACTGGACAGACTGATGATCACTGGAGAACCTCTGAGACCAAGTCAATCCAGACAG 661  
DB 9072 CACTGGACAGACTGATGATCACTGGAGAACCTCTGAGACCAAGTCAATCCAGACAG 9131  
QY 662 ACAGATGTGACAGAGCAAAAGTGCATATATGCAATATGTTAAATGTGATTTACAC 721  
DB 9132 ACAGATGTGACAGAGCAAAAGTGCATATATGCAATATGTTAAATGTGATTTACAC 9191

OY		722	CTAGCTATGGGACATCGTGGCTCCCTTAAGTCACAGAAATCATGGGGGTATGACTGCCTCCAA	781
Dd		9192	CTAGCTATGGGACTCTCTGGCTCTTAGTCACAGAAATCATGGGGGTATGACTGCCTCCAA	9251
OY		782	CCTGTGGGCTGTATGAACAAGCTCAGGCTAGTCTCCCCAAGTGGGGGCTGTGCCCTCCCTG	841
Dd		9252	CCCTGTGGGCTGTATGAACAAGCTCAGGCTAGTCTCCCCAAGTGGGGGCTGTGCCCTCCCTG	9310
OY		842	GGAAGGTTCCGTGGGACAGCCCATTACATGTGTTCATATGTGTGAATAATGATGCTAAACC	901
Dd		9311	GGACGGTTCCGTGGGACAGCCCATTACATGTGTTCATATGTGTGAATAATGATGCTAAAGCC	9370
OY		902	CCTGTGCTGCTGCTGCATGCAATGCCACACAGAGCGGTGGGGGCTGGTGGGGACAATCCAT	961
Dd		9371	CCTGTGCTGCTGCTGCATGCAATGCCACACAGAGCGGTGGGGGCTGGTGGGGACAATCCAT	9430
OY		962	CGTGAAGTGTGTCTCTCAGCTTAGTAGTCTTGAGACAGAGACTTGGCGGGGGATGCTCCAGAT	1021
Dd		9431	CGTGAAGTGTGTCTCTCAGCTTAGTAGTCTTGAGACAGAGACTTGGCGGGGGATGCTCCAGAT	9490
OY		1022	GTGGGTGATTTCTGTACCTGGGAGGCTATCTGTGACCTCCGACAGGGGACACTGCCAGG	1089
Dd		9491	GTGGGTGATTTCTGTACCTGGGAGGCTATCTGTGACCTCCGACAGGGGACACTGCCAGG	9550
OY		1082	CCAGCCACAGGGGTCAAGGGGCAAGGGTGCACACTCAGCATGAGCCAGAATGGGGTCAAG	1141
Dd		9551	CCAGCCACAGGGGTCAAGGGGCAAGGGTGCACACTCAGCATGAGCCAGAATGGGGTCAAG	9610
OY		1142	GAGCAGGTGTGGTTTGTAGCCAGAGACTGGGGGCGGGGTGTGGGGCCGGGCTTTCTGCTC	1201
Dd		9611	GAGCAGGTGTGGTTTGTAGCCAGAGACTGGGGGCGGGGTGTGGGGCCGGGCTTTCTGCTC	9670
OY		1202	AATTGCTTCATGAAAGCTCAAAAGCACAGCAAACAGGGCTTCCCTGCTCGAGTT	1261
Dd		9671	AATTGCTTCATGAAAGCTCAAAAGCACAGCAAACAGGGCTTCCCTGCTCGAGTT	9730
OY		1262	TGAATATCCAGAACCTTTTGTACTCTTGTGGTAAATGTTTATTTTTGTAAAAATA	1321
Dd		9731	TGAATATCCAGAACCTTTTGTACTCTTGTGGTAAATGTTTATTTTTGTAAAAATA	9790
OY		1322	AAATTAATTTAGTTAATTAATAATGATGTTTCAACAGCAAACTCTTCCCTTA	1370
Dd		9791	AAATTAATTTAGTTAATTAATAATGATGTTTCAACAGCAAACTCTTCCCTTA	9839
<hr/>				
RESULT 3				
ID AAH98172 standard; cDNA, 1846 BP.				
XX	AAH98172;			
XX				
DT	12-OCT-2001 (first entry)			
DE	Human EST-derived coding sequence SEQ ID NO: 29.			
KW	Human; sheep; pig; cow; fruit fly; yeast; hamster; macaque; horse;			
KM	tomato; monkey; dog; sea urchin; expressed sequence tag; EST;			
KX	diagnostics; forensic test; gene mapping; genetic disorder;			
KM	biodiversity; gene therapy; nutrition; ss.			
OS	Homo sapiens.			
PN	WO200154477-A2.			
PD	02-AUG-2001.			
PF	25-JAN-2001; 2001WO-US02687.			
PR	25-JAN-2000; 2000US-0491404.			
PR	17-JUL-2000; 2000US-0617746.			
PR	03-AUG-2000; 2000US-0631451.			
PR	15-SEP-2000; 2000US-063870.			



Db 2435 ATGATGTTTCACGACAACTCTTCCT 2461

# RESULT 5

ABN94640/c ID ABN94640 standard; DNA; 427 BP.

AC ABN94640;

DT 13-AUG-2002 (first entry)

DE Gene #1138 used to diagnose liver cancer.

XX Gene; liver cancer; db; hepatocellular carcinoma; hepatocytic;  
 KW metastatic liver tumor; cytotoxic; expression profile; disease state;  
 KM disease progression; drug toxicity; drug efficacy; drug metabolism.

OS Homo sapiens.

PN WO200229103-A2.

PD 11-APR-2002.

PF 02-OCT-2001; 2001MO-US30589.

PR 02-OCT-2000; 2000US-237054P.

PA (GENE-) GENE LOGIC INC.

PI Horne D, Alvares C, Peres-Da-Silva S, Vockley JG;

DR WPI; 2002-426119/45.

XX Diagnosing and detecting the progression of liver cancer,

PT hepatocellular carcinoma or metastatic liver tumor in a patient,

PT involves detecting the level of expression of two or more genes in a

PT liver tissue sample

XX Claim 1; SEQ ID NO 1138; 298bp; English.

XX The invention relates to a novel method for diagnosing and detecting the

CC progression of liver cancer, hepatocellular carcinoma or metastatic liver

CC tumour in a patient, and differentiating metastatic liver cancer from

CC hepatocellular carcinoma in a patient, involving detecting the level of

CC expression of two or more genes represented in ABN93503-ABN97455 in a

CC tissue sample. The method of the invention has hepatocytic, and

CC cytosolic activity. The method is useful for diagnosing and detecting

CC the progression of liver cancer, hepatocellular carcinoma and metastatic

CC liver carcinoma in a patient. The method is useful for identifying

CC expression profiles which serve as useful diagnostic markers as well as

CC markers that can be used to monitor disease states, disease progression,

CC drug toxicity, drug efficacy and drug metabolism.

CC Note: The sequence data for this patent did not form part of the printed

CC specification, but was obtained in electronic format directly from WIPO

CC at ftp.wipo.int/pub/published\_pct\_sequences.

XX Sequence 427 BP; 107 A; 126 C; 92 G; 102 T; 0 other;

XX Query Match 23.0%; Score 322; DB 24; Length 427;

XX Best Local Similarity 99.5%; Pred. No. 2.4e-115;

XX Matches 422; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 936 GTGGGGGCTGCTGGGGGCAATTCATCGTGGAGTCTTCTCTACCTTAGTCTGGACAG 995

DB 427 GTGGGGGCTGCTGGGGGCAATTCATCGTGGAGTCTTCTCTACCTTAGTCTGGACAG 368

QY 996 AGACTTGGCGGGGATGCTCCAGATGTTGGTATTTCTGTACCTGGGAGGCTATCTCTG 1055

DB 367 AGACTTGGCGGGGATGCTCCAGATGTTGGTATTTCTGTACCTGGGAGGCTATCTCTG 308

QY 1056 ACCGCCGACAGGGGACACTCCAGGCCAGCCAGGGGTGAGGGGACAGAGTGACACT 1115

Db 307 ACTCCGACAGGGGACACTCCAGGCCAGCCAGGGGTGAGGGGACAGAGTGACACT 248

QY 1116 CAGCATGAGCCAAAGACTGGGGTCCAGGACAGGAGTGGTTGAGCCAGGACTGGGGCGG 1175

Db 247 CAGCATGAGCCAAAGACTGGGGTCCAGGACAGGAGTGGTTGAGCCAGGACTGGGGCGG 188

QY 1176 GGGTGGGCGCGGGGCTTTTCGCTTCATTTGCTTTTCAATGAAAGCTTCMAAGCCCAA 1235

Db 187 GGGTGGGACCGGGGCTTTTCGCTTCATTTGCTTTTCAATGAAAGCTTCMAAGCCCAA 128

QY 1236 ACCAGGCTTTCCCTTCCTCGAGTTGAATTCAGAAATCTTTTGTACTCTTGTGCT 1295

Db 127 ACCAGGCTTTCCCTTCCTCGAGTTGAATTCAGAAATCTTTTGTACTCTTGTGCT 68

QY 1236 TAAATGTTTATTTTGTAAATAAATAAATAAATAAATAAATAAATAAATAAATAA 1355

Db 67 TAAATGTTTATTTTGTAAATAAATAAATAAATAAATAAATAAATAAATAAATAA 8

QY 1356 CAAA 1359

Db 7 CAAA 4

# RESULT 6

AAH86413/c ID AAH86413 standard; DNA; 150 BP.

AC AAH86413;

DT 27-FEB-2002 (first entry)

DE Human single nucleotide polymorphism containing DNA sequence #1270.

XX Biallelic marker; polymorphism; human; disease; diagnosis; treatment;

KW phenotypic trait; gene therapy; forensic; paternity; mapping; cancer;

KW transgenic; single nucleotide polymorphism; SNP; ss.

OS Homo sapiens.

XX Key Location/Qualifiers

FT Variation replace(25,T) /tag= a

FT /standard\_name= "single nucleotide polymorphism"

PN MO9953095-A2.

XX 21-OCT-1999.

PD 30-MAR-1999; 99WO-US06893.

PF 09-APR-1998; 98US-0057871.

PR (WHED) WHITEHEAD INST BIOMEDICAL RES.

PA Lander ES, Wang D, Hudson T;

PI WPI; 1999-620443/53.

DR Polymorphic human genomic sequences and related allele-specific probes

XX of disease

XX Claim 1; Page 167; 330pp; English.

XX This invention describes novel human nucleic acid segments (I)

XX containing polymorphic sites. The polymucleotides of (I) are used for,

XX e.g. correlating disease polymorphisms (or disease susceptibility) or

XX other phenotypic traits (e.g. baldness, obesity, fertility, strength,

XX response to drugs etc.); diagnosing and monitoring e.g. cancer,

XX inflammation, heart or central nervous system diseases; detecting

XX susceptibility to microbial infection; treating or preventing such

XX diseases; forensic analysis; gene therapy; paternity testing; mapping

XX genomic loci associated with phenotypic traits (and subsequent cloning



CC of the genes responsible); and the production of transgenic organisms.  
CC Antibodies raised against (I) are useful as diagnostic and therapeutic  
CC tools and in drug screening. AAH85144 - AAH87644 represent the human  
CC DNA sequences containing biallelic polymorphic sites described in the  
CC invention.

XX Sequence 150 BP; 52 A; 19 C; 30 G; 48 T; 1 other;

Query Match 8.1%; Score 113; DB 20; Length 150;  
Best Local Similarity 100.0%; Pred. No. 2.8e-34;  
Matches 113; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1256 CGAGTTGGAATATCCAGAACTCTTTGACTCTTGTGGTTAAATGTTATTTTGTA 1315

DB 113 CGAGTTGGAATATCCAGAACTCTTTGACTCTTGTGGTTAAATGTTATTTTGTA 54

QY 1316 AAAATTAATAATTAAGTTAATAAATGATGTTTACAGCAAACTCTTCCT 1368

DB 53 AAAATTAATAATTAAGTTAATAAATGATGTTTACAGCAAACTCTTCCT 1

RESULT 7

AAH86414/C  
ID AAH86414 standard; DNA; 150 BP.

XX AAH86414;

DT 27-FEB-2002 (first entry)

DE Human single nucleotide polymorphism containing DNA sequence #1271.

XX Biallelic marker; polymorphism; human; disease; diagnosis; treatment;

KW phenotypic trait; gene therapy; forensic; paternity; mapping; cancer;

KW transgenic; single nucleotide polymorphism; SNP; ss.

XX Homo sapiens.

OS Location/Qualifiers

Key replacement(92,T) /tag= a

FT Variation /standard\_name= "single nucleotide polymorphism"

XX WO953095-A2.

XX 21-OCT-1999.

XX 30-MAR-1999; 99WO-US06893.

XX 09-APR-1998; 98US-0057871.

XX (WHED) WHITEHEAD INST BIOMEDICAL RES.

XX Lander ES, Wang D, Hudson T;

XX WPI; 1999-620443/53.

PT polymorphic human genomic sequences and related allele-specific probes

PT of disease - useful for genetic analysis, e.g. diagnosis and monitoring

XX Claim 1; Page 167; 330pp; English.

CC This invention describes novel human nucleic acid segments (I) containing polymorphic sites. The polymorphisms of (I) are used for, e.g. correlating disease polymorphisms (or disease susceptibility) or other phenotypic traits (e.g. baldness, obesity, fertility, strength, response to drugs etc.); diagnosing and monitoring e.g. cancer, inflammation, heart or central nervous system diseases; detecting susceptibility to microbial infection; treating or preventing such diseases; forensic analysis; gene therapy; paternity testing; mapping genomic loci associated with phenotypic traits and subsequent cloning of the genes responsible; and the production of transgenic organisms. Antibodies raised against (I) are useful as diagnostic and therapeutic

CC tools and in drug screening. AAH85144 - AAH87644 represent the human  
CC DNA sequences containing biallelic polymorphic sites described in the  
CC invention.

XX Sequence 150 BP; 52 A; 19 C; 30 G; 48 T; 1 other;

Query Match 8.1%; Score 113; DB 20; Length 150;  
Best Local Similarity 100.0%; Pred. No. 2.8e-34;  
Matches 113; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1256 CGAGTTGGAATATCCAGAACTCTTTGACTCTTGTGGTTAAATGTTATTTTGTA 1315

DB 113 CGAGTTGGAATATCCAGAACTCTTTGACTCTTGTGGTTAAATGTTATTTTGTA 54

QY 1316 AAAATTAATAATTAAGTTAATAAATGATGTTTACAGCAAACTCTTCCT 1368

DB 53 AAAATTAATAATTAAGTTAATAAATGATGTTTACAGCAAACTCTTCCT 1

RESULT 8

AAH87838  
ID AAH87838 standard; cDNA; 1837 BP.

XX AAH87838;

DT 13-FEB-2002 (first entry)

DE DNA encoding novel human diagnostic protein #14642.

XX Human; chromosome mapping; gene mapping; gene therapy; forensic;

KW food supplement; medical imaging; diagnostic; genetic disorder; ss.

XX Homo sapiens.

XX WO200175067-A2.

XX 11-OCT-2001.

XX 30-MAR-2001; 2001WO-US08631.

XX 31-MAR-2000; 2000US-0540217.

XX 23-AUG-2000; 2000US-0649167.

XX (HYSE-) HYSEQ INC.

XX Drmanac RT, Liu C, Tang YT;

XX WPI; 2001-639362/73.

XX P-PSDB; ABG14651.

PT New isolated polynucleotide and encoded polypeptides, useful in

PT diagnostics, forensics, gene mapping, identification of mutations

PT responsible for genetic disorders or other traits and to assess

XX Claim 1; SEQ ID NO 14642; 103pp; English.

CC The invention relates to isolated polynucleotide (I) and  
CC polypeptide (II) sequences. (I) is useful as hybridisation probes,  
CC polymerase chain reaction (PCR) primers, oligomers, and for chromosome  
CC and gene mapping, and in recombinant production of (II). The  
CC polynucleotides are also used in diagnostics as expressed sequence tags  
CC for identifying expressed genes. (I) is useful in gene therapy techniques  
CC to restore normal activity of (II) or to treat disease states involving  
CC (II). (II) is useful for generating antibodies against it, detecting or  
CC quantitating a polypeptide in tissue, as molecular weight markers and as  
CC a food supplement. (II) and its binding partners are useful in medical  
CC imaging of sites expressing (II). (I) and (II) are useful for treating  
CC disorders involving aberrant protein expression or biological activity.  
CC The polypeptide and polynucleotide sequences have applications in  
CC diagnostics, forensics, gene mapping, identification of mutations  
CC responsible for genetic disorders or other traits to assess biodiversity  
CC and to produce other types of data and products dependent on DNA and



CC amino acid sequences. AAs64197-AAs94564 represent novel human  
CC diagnostic coding sequences of the invention.  
CC Note: The sequence data for this patent did not appear in the printed  
CC specification, but was obtained in electronic format directly from WIPO  
CC at ftp.wipo.int/pub/published\_pct\_sequences.

CC Sequence 1837 BP; 424 A; 540 C; 471 G; 402 T; 0 other;

Query Match 4.9%; Score 69; DB 23; Length 1837;  
Best Local Similarity 100.0%; Pred. No. 2.4e-17;  
Matches 69; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 28 GTATCTGCAAGACTCCACGCGCAAGCAGAGTACCGAGTACCGATGATCCAGACGAC 87  
DB 1033 GTATCTGCAAGACTCCACGCGCAAGCAGAGTACCGAGTACCGATGATCCAGACGAC 1092

QY 88 CATCCCCCA 96  
DB 1093 CATCCCCCA 1101

RESULT 9  
AAs78836  
ID AAs78836 standard; cDNA; 370 BP.

AC AAs78836;

DT 13-FEB-2002 (first entry)

DE DNA encoding novel human diagnostic protein #14640.

KW Human; chromosome mapping; gene mapping; gene therapy; forensic;  
KW food supplement; medical imaging; diagnostic; genetic disorder; ss.

OS Homo sapiens.

PN WO200175067-A2.

XX 11-OCT-2001.

PF 30-MAR-2001; 2001MO-US08631.

PR 31-MAR-2000; 2000US-0540217.

PR 23-AUG-2000; 2000US-0649167.

PA (HYSR-) HYSRQ INC.

PI Drmanac RT, Liu C, Tang YT;

DR WPI: 2001-639362/73.

DR P-PSDB; AAs714649.

PT New isolated polynucleotide and encoded polypeptides, useful in  
PT diagnostics, forensics, gene mapping, identification of mutations  
PT responsible for genetic disorders or other traits and to assess  
PT biodiversity

PS Claim 1; SEQ ID No 14640; 103bp; English.

XX The invention relates to isolated polynucleotide (I) and  
CC polypeptide (II) sequences. (I) is useful as hybridisation probes,  
CC polymerase chain reaction (PCR) primers, oligomers, and for chromosome  
CC and gene mapping, and in recombinant production of (II). The  
CC polynucleotides are also used in diagnostics as expressed sequence tags  
CC for identifying expressed genes. (I) is useful in gene therapy techniques  
CC to restore normal activity of (II) or to treat disease states involving  
CC (II). (II) is useful for generating antibodies against it, detecting or  
CC quantitating a polypeptide in tissue, as molecular weight markers and as  
CC a food supplement. (II) and its binding partners are useful in medical  
CC imaging of sites expressing (II). (I) and (II) are useful for treating  
CC disorders involving aberrant protein expression or biological activity.  
CC The polypeptide and polynucleotide sequences have applications in  
CC diagnostics, forensics, gene mapping, identification of mutations

CC responsible for genetic disorders or other traits to assess biodiversity  
CC and to produce other types of data and products dependent on DNA and  
CC amino acid sequences. AAs64197-AAs94564 represent novel human  
CC diagnostic coding sequences of the invention.  
CC Note: The sequence data for this patent did not appear in the printed  
CC specification, but was obtained in electronic format directly from WIPO  
CC at ftp.wipo.int/pub/published\_pct\_sequences.

CC Sequence 370 BP; 67 A; 126 C; 101 G; 76 T; 0 other;

Query Match 3.8%; Score 53; DB 23; Length 370;  
Best Local Similarity 100.0%; Pred. No. 4.5e-11;  
Matches 53; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 95 CAGGAAGACTACCGCTGCTGCTGCTACCTACCAAGGAGGAGTGCCTTC 147  
DB 317 CAGGAAGACTACCGCTGCTGCTGCTACCTACCAAGGAGGAGTGCCTTC 369

RESULT 10  
AAs44726  
ID AAs44726 standard; cDNA; 1886 BP.

AC AAs44726;

DT 27-MAR-2001 (first entry)

DE Novel protein kinase cDNA, SEQ ID NO: 107.

KW Human; mouse; protein kinase; antiarthritic; antisclerotic; osteopathic;  
KW immunosuppressive; cardiac; renal; antiinflammatory; antiaesthetic;  
KW dermatological; antidiabetic; antifertility; gene therapy; vaccine;  
KW immune disorder; cardiovascular disease; neurodegenerative disease;  
KW cancer; autoimmune disorder; stroke; inflammatory bowel disease;  
KW inflammatory pelvic disease; multiple sclerosis; psoriasis; ss.

OS Mus musculus.

PN WO200073469-A2.

XX 07-DEC-2000.

PF 26-MAY-2000; 2000MO-US14842.

PR 28-MAY-1999; 99US-0136503.

PA (SUGR-) SUGEN INC.

PI Plowman GD, Martinez R, Whyte D, Sudersanam S;

DR WPI: 2001-032161/04.

DR P-PSDB; AAs59576.

PT Nucleic acids encoding kinase polypeptides, useful for diagnosing and  
PT treating immune-related diseases and disorders, cardiovascular disease,  
PT neurodegenerative diseases and/or cancers

PS Disclosure; Fig 2; 310bp; English.

XX The present sequence encodes a novel protein kinase. The nucleic acids  
CC and the protein kinases they encode may be used in the treatment and  
CC diagnosis of diseases associated with inappropriate kinase expression  
CC such as immune-related diseases and disorders, cardiovascular disease,  
CC neurodegenerative diseases and/or cancers. The nucleic acids and  
CC complementary sequences may also be used as DNA probes in diagnostic  
CC assays. The kinase polypeptides may be used as antigens in the production  
CC of antibodies of kinase expression and activity. Anti-kinase antibodies  
CC and kinase antagonists may also be used to down regulate kinase  
CC expression and activity. Diseases related to kinase expression and  
CC activity include Rheumatoid arthritis, atherosclerosis, autoimmune  
CC disorders, complications of organ transplantation, myocardial infarction,  
CC immune disorders, cardiomyopathies, strokes, renal failure,  
CC oxidative-stress related disorders, chronic inflammatory bowel disease,

CC chronic inflammatory pelvic disease, multiple sclerosis, asthma,  
CC osteoarthritis, psoriasis, rhinitis, autoimmunity, diabetes, cancers and  
CC reproductive disorders.

XX Sequence 1886 BP; 384 A; 559 C; 553 G; 390 T; 0 other;

Query Match 2.6%; Score 36; DB 22; Length 1886;

Best Local Similarity 100.0%; Pred. No. 0.00015;

Matches 36; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 203 GCCTTGTGTGTCACCAACGACCTGACAGGT 238

DB 981 GCCTTGTGTGTCACCAACGACCTGACAGGT 1016

RESULT 11

AB085200/c

ID AB085200 standard; DNA; 386 BP.

AC AB085200;

DT 05-SEP-2002 (first entry)

DE Arabidopsis thaliana expressed polynucleotide SEQ ID NO 70.

XX Plant; Arabidopsis; transgenic; fungicide; insecticide; ds.

XX Arabidopsis thaliana.

PN US2002062014-A1.

PD 23-MAY-2002.

PF 26-JUN-2001; 2001US-0770791.

PR 27-JUN-2000; 2000US-178480P.

XX (GORL/) GORLACH J.

PA (ANY/) AN Y.

PA (HAM/) HAMILTON C M.

PA (PRIC/) PRICE J L.

PA (RAIN/) RAINES T M.

PA (YUYV/) YU Y.

PA (RAME/) RAMEAKA J G.

PA (PAGE/) PAGE A.

PA (MATH/) MATHW A V.

PA (LEDF/) LEDFORD B L.

PA (WOES/) WOESSNER J P.

PA (HAAS/) HAAS W D.

PA (GARC/) GARCIA C A.

PA (KRIC/) KRICKER M.

PA (SLAT/) SLATER T.

PA (DAVI/) DAVIS K R.

PA (ALLE/) ALLEN K.

PA (HOFF/) HOFFMAN N.

PA (HURB/) HURBAN P.

XX

PI Gorlach J, An Y, Hamilton CM, Price JL, Raines TM, Yu Y,

PI Rameka JG, Page A, Mathew AV, Ledford BL, Woessner JP, Haas WD,

PI Garcia CA, Kricker M, Slater T, Davis KR, Allen K, Hoffman N,

PI Hurban P;

XX

DR WPI; 2002-479265/51.

XX

PT New nucleic acid sequences of Arabidopsis thaliana and their encoded

PT products are useful to produce transgenic plants, to screen for

PT biologically active agents such as fungicides and insecticides and in

PT genetic studies

XX

PS Claim 1; SEQ ID NO 70; 1886 + Sequence Listing; English.

CC The invention relates to a novel nucleic acid of Arabidopsis thaliana

CC comprising a sequence capable of hybridising under stringency to one of

CC the 999 sequences referred to but not defined in the specification  
CC (AB085131-AB086129). The nucleic acid sequences are useful to identify  
CC homologous or related genes, to produce compositions that modulate  
CC expression or function of the encoded protein, to map functional regions  
CC of the protein, to study associated physiological pathways, to  
CC genetically manipulate cells and plants. The encoded products are useful  
CC to screen for biologically active agents such as fungicides or  
CC insecticides and to elucidate biochemical pathways.

XX Sequence 386 BP; 116 A; 76 C; 75 G; 119 T; 0 other;

Query Match 2.1%; Score 30; DB 24; Length 386;

Best Local Similarity 100.0%; Pred. No. 0.037;

Matches 30; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1361 TCTCCCTTAAAAA 1390

DB 31 TCTCCCTTAAAAA 2

RESULT 12

ABL80663/c

ID ABL80663 standard; cDNA; 421 BP.

AC ABL80663;

DT 17-MAY-2002 (first entry)

DE Human ovarian cancer related cDNA clone SEQ ID NO:3641.

XX Human; ovarian cancer; ovarian tumour; cytostatic; gene; ss.

XX Homo sapiens.

PN WO200192581-A2.

PD 06-DEC-2001.

PF 29-MAY-2001; 2001WO-US17756.

PR 26-MAY-2000; 2000US-207484P.

XX (CORI-) CORIXA CORP.

PA Algate PA, Harlocker SL, Jones R;

XX WPI; 2002-122075/16.

DR

XX

PT Composition for therapy and diagnosis of ovarian cancer comprising

PT polypeptide of a ovarian tumor polypeptide, polynucleotide encoding

PT polypeptide, antibody specific to polypeptide or T cell expressing

PT polypeptide

XX

PS Claim 1; SEQ ID 3641; 489bp; English.

XX

CC The present invention describes a composition (I) comprising: carriers

CC and immunostimulants; and a polypeptide (II) of a ovarian tumour

CC polypeptide encoded by a polynucleotide (III) having a cDNA sequence

CC (S1) from the 10912 nucleotide sequences as given in ABL77023 to

CC ABL87934, (III) encoding (II) having a sequence (S2), a T cell

CC population of (II), or antigen presenting cells that express (II).

CC (S1) has cytostatic activity. An oligonucleotide (IV) that hybridises to

CC (S1) can be used for detecting ovarian cancer in a patient's biological

CC sample preferably serum or ovarian tissue. The method comprises

CC contacting a biological sample from a patient with (IV), detecting the

CC amount of polynucleotide hybridising to (IV) and comparing the amount to

CC a predetermined cutoff value and thereby detecting ovarian cancer in the

CC patient, where the amount of polynucleotide hybridising to (IV) is

CC detected preferably by polymerase chain reaction (PCR). (I) comprising

CC (III) and/or (II) is useful for stimulating and/or expanding T cells

CC specific for an ovarian tumour protein comprising contacting T cells

CC with (III) or (II). (III) is useful in design and preparation of

CC ribozyme molecules for inhibiting expression of the tumour polypeptides

CC and proteins in tumour cells; and to isolate a full length gene from a  
CC suitable library e.g., a tumour cDNA library using well known  
CC techniques.

XX Sequence 421 BP; 113 A; 77 C; 100 G; 131 T; 0 other;

XX Query Match 2.1%; Score 30; DB 24; Length 421;  
XX Best Local Similarity 100.0%; Pred. No. 0.037;  
XX Matches 30; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1361 TCTTCCTTAAAAA 1390  
Db 30 TCTTCCTTAAAAA 1

RESULT 13  
ABN99116/c  
ID ABN99116 standard; DNA; 777 BP.

AC ABN99116;

DT 01-AUG-2002 (first entry)

DE Arabidopsis thaliana expressed polynucleotide SEQ ID NO 884.

XX Arabidopsis thaliana, plant; insecticide; fungicide; transgenic; stress;  
XX disease; crop; thale cress; tolerance factor; insect; pathogen;  
XX nutrition; ds.

OS Arabidopsis thaliana.

PN US2002023281-A1.

PD 21-FEB-2002.

PF 26-JAN-2001; 2001US-0770445.

PR 27-JAN-2000; 2000US-178472P.

XX (GORL/) GORLACH J.

XX (ANYV/) AN Y.

XX (HAMT/) HAMILTON C M.

XX (PRIC/) PRICE J L.

XX (RAIN/) RAINES T M.

XX (YUYV/) YU Y.

XX (RAME/) RAMEAKA J G.

XX (PAGE/) PAGE A.

XX (MATH/) MATHW A V.

XX (LEDF/) LEDFORD B L.

XX (WOES/) WOESSNER J P.

XX (HAAS/) HAAS W D.

XX (GARC/) GARCIA C A.

XX (KRICK/) KRICKER M.

XX (SLAT/) SLATER T.

XX (DAVI/) DAVIS K R.

XX (ALIE/) ALLEN K.

XX (HOFF/) HOFFMAN N.

XX (HURE/) HUREN P.

XX GORLACH J, An Y, Hamilton CM, Price JL, Raines TM, Yu Y,

PI Ramaea JG, Page A, Mathew AV, Ledford BL, Woessner JP, Haas WD;

PI Garcia CA, Kricker M, Slater T, Davis KR, Allen K, Hoffman N;

PI Hurdan P;

XX WPI; 2002-400781/43.

XX New Arabidopsis thaliana nucleic acid for identifying homologous genes,

PT producing compositions that modulate the expression or function of its

PT encoded protein, and mapping functional regions of protein

XX Claim 1; SEQ ID NO 884; 49bp + Sequence Listing; English.

XX The invention relates to an Arabidopsis thaliana nucleic acid (I)

CC comprising a sequence capable of hybridising under stringent conditions  
CC to a sequence selected from any one of 999 sequences (ABN98233-ABN99231),  
CC given in the specification or its fragment. A polypeptide (II) encoded by

CC (I), a transgenic plant (III) comprising an exogenous nucleic acid or a

CC genetically modified cell (IV) comprising an exogenous nucleic acid, is

CC useful for screening a candidate agent for its biological effect. (I) is

CC useful in identifying homologous or related genes, in producing

CC compositions that modulate the expression or function of its encoded

CC protein, mapping functional regions of the protein and in studying

CC associated physiological pathways. (I) is also useful for the genetic

CC manipulation of cells, particularly plant cells. (I) is also useful in

CC screening assays of various plant strains to determine the strains that

CC are best capable of withstanding a particular disease or environmental

CC stress. (II) and (III) are useful for screening of biologically active

CC agents, e.g. fungicides, insecticides, etc., for elucidating biochemical

CC pathways. The screened agents are useful in improved methods of treating

CC crops to prevent or treat disease. (II) are also useful in screening

CC programs to identify agents that mimic or enhance the action of tolerance

CC factors. Such agents are useful in improved methods of treating crops to

CC enhance their tolerance to environmental stress. (I) is also useful

CC for enhancing or inhibiting production of a biosynthetic product in a

CC plant. (III) is useful for identifying other mediators that may induce

CC expression of proteins of interest, for establishing the extent to which

CC any specific insect and/or pathogen is responsible for damage to a

CC particular plant, for identifying other mediators that enhance or induce

CC tolerance to environmental stress, for identifying factors involved in

CC biosynthetic pathways of nutritional, commercial, or medicinal value and

CC for identifying products of nutritional, commercial or medicinal

CC value. (IV) is useful in the study of genetic function and regulation,

CC for alteration of the cellular metabolism and for screening compounds

CC that may affect the biological function of the gene or gene products.

CC Note: The sequence data for this patent did not form part of the printed

CC specification, but was obtained in electronic format directly from USPRO

CC at [seqdata.uspto.gov/sequence.html?DocID=999909770445](http://seqdata.uspto.gov/sequence.html?DocID=999909770445).

XX SQ Sequence 777 BP; 192 A; 196 C; 149 G; 240 T; 0 other;

XX Query Match 2.1%; Score 30; DB 24; Length 777;

XX Best Local Similarity 100.0%; Pred. No. 0.034;

XX Matches 30; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1361 TCTTCCTTAAAAA 1390

Db 45 TCTTCCTTAAAAA 16

RESULT 14

AAS46209

ID AAS46209 standard; cDNA; 2457 BP.

XX AAS46209;

XX 18-DEC-2001 (first entry)

XX Human DNA encoding PRO polypeptide sequence #285.

XX PRO polypeptide; mammal; tumour; cancer; human; cattle; horse; sheep; ss;

XX dog; cat; pig; goat; rabbit; tumour necrosis factor alpha; TNF-alpha;

XX blood; chondrocyte cell; cell proliferation; cell differentiation; colon;

XX adrenal; lung; breast; prostate; rectum; cervix; liver; genetic disorder;

XX PCR primer.

XX Homo sapiens.

XX WO200168848-A2.

XX 20-SEP-2001.

XX 28-FEB-2001; 2001WO-US06520.

XX 01-MAR-2000; 2000WO-US05601.

XX 02-MAR-2000; 2000WO-US05841.

XX 03-MAR-2000; 2000US-187202P.

PR 06-MAR-2000; 2000US-186968P.  
 PR 14-MAR-2000; 2000US-189320P.  
 PR 14-MAR-2000; 2000US-189328P.  
 PR 15-MAR-2000; 2000MO-US06884.  
 PR 21-MAR-2000; 2000US-190828P.  
 PR 21-MAR-2000; 2000US-191007P.  
 PR 21-MAR-2000; 2000US-191048P.  
 PR 21-MAR-2000; 2000US-191314P.  
 PR 28-MAR-2000; 2000US-192655P.  
 PR 29-MAR-2000; 2000US-193032P.  
 PR 29-MAR-2000; 2000US-193053P.  
 PR 30-MAR-2000; 2000MO-US08439.  
 PR 04-APR-2000; 2000US-194449P.  
 PR 04-APR-2000; 2000US-194647P.  
 PR 11-APR-2000; 2000US-195975P.  
 PR 11-APR-2000; 2000US-196000P.  
 PR 11-APR-2000; 2000US-196187P.  
 PR 11-APR-2000; 2000US-196690P.  
 PR 11-APR-2000; 2000US-196820P.  
 PR 18-APR-2000; 2000US-198121P.  
 PR 18-APR-2000; 2000US-198585P.  
 PR 25-APR-2000; 2000US-199397P.  
 PR 25-APR-2000; 2000US-199550P.  
 PR 25-APR-2000; 2000US-199549P.  
 PR 03-MAY-2000; 2000US-201516P.  
 PR 17-MAY-2000; 2000MO-US13705.  
 PR 22-MAY-2000; 2000MO-US14042.  
 PR 30-MAY-2000; 2000MO-US14941.  
 PR 02-JUN-2000; 2000MO-US15264.  
 PR 05-JUN-2000; 2000US-209812P.  
 PR 28-JUL-2000; 2000MO-US20710.  
 PR 22-AUG-2000; 2000US-0644848.  
 PR 24-AUG-2000; 2000MO-US23328.  
 PR 08-NOV-2000; 2000MO-US30952.  
 PR 01-DEC-2000; 2000MO-US32578.  
 PR 20-DEC-2000; 2000MO-US34556.

## (GETH ) GENENTECH INC.

Baker KP, Chen J, Desnoyers L, Goddard A, Godowski PJ, Gurney AL,  
 Pan J, Smith V, Watanabe CK, Wood WT, Zhang Z;  
 P-PSDB; AA029308.

Novel nucleic acids encoding PRO polypeptides, used to diagnose the  
 presence of tumours, such as prostate and breast tumours, in mammals and  
 to screen for modulators of the compounds -

Claim 2; Fig 569; 774pp; English.

CC Sequences AAS45925-AAS46231 represent DNA molecules encoding and PCR  
 CC primers for PRO polypeptides of the invention. The sequences of the  
 CC invention can be used to detect the presence of a tumour in a mammal by  
 CC comparing the level of expression of a PRO polypeptide in a test sample  
 CC of cells from the animal and a control sample of normal cells, whereby a  
 CC higher level of expression in the test sample indicates the presence of a  
 CC tumour in the mammal. Mammals include dogs, cats, cattle, horses, sheep,  
 CC pigs, goats and rabbits but are preferably human. The polypeptides can be  
 CC used to stimulate tumour necrosis factor (TNF) alpha release from human  
 CC blood, when contacted with it. A specific polypeptide can be used to  
 CC stimulate the proliferation or differentiation of chondrocyte cells. The  
 CC PRO proteins can be used to determine the presence of tumours and also  
 CC susceptibility to tumour development, particularly adrenal, lung, colon,  
 CC breast, prostate, rectal, cervical, or liver tumours, in mammalian  
 CC subjects. The oligonucleotide probes specific for the PRO nucleic acids  
 CC can be used for genetic analysis of individuals with genetic disorders.

XX Sequence 2457 BP; 678 A; 629 C; 568 G; 582 T; 0 other;

## Query Match

Best Local Similarity 2.1%; Score 30; DB 22; Length 2457;  
 Matches 30; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1361 TCTTCCTAAAAAAAAAAAAAAAAAAAA 1390  
 Db 2423 TCTTCCTAAAAAAAAAAAAAAAAAAAA 2452

RESULT 15  
 AAD31150/c  
 ID AAD31150 standard; cDNA; 3239 BP.

XX AAD31150;

DT 27-AUG-2002 (first entry)

DE Human carbohydrate-associated protein (CARBAP)-2 cDNA.

XX Human; carbohydrate-associated protein-2; CARBAP-2; arteriosclerosis;  
 KW cell proliferative disorder; cancer; autoimmune; inflammatory disorder;  
 KW allergy; anaemia; asthma; infection; reproductive disorder; infertility;  
 KW ovulatory defect; neurological disorder; Alzheimer's disease;  
 KW Parkinson's disease; gene therapy; ss.

XX Homo sapiens.

OS Key Location/Qualifiers

FH CDS 1969..2439

FT /tag= a

FT /product= a "Carbohydrate-associated protein (CARBAP)-2"

PN WO200229055-A2.

PD 11-APR-2002.

PF 01-OCT-2001; 2001MO-US30591.

PR 02-OCT-2000; 2000US-237456P.

XX (INCY-) INCYTE GENOMICS INC.

PI Tang YT, Elliott VS, Yue H, Lal P, Malia NK;

DR WPI; 2002-426114/45.

XX P-PSDB; AAE14720.

PT Novel human carbohydrate-associated proteins and genes, useful in the

PT diagnosis, prevention and treatment of cell proliferative,  
 PT autoimmune/inflammatory, reproductive and neurological disorders -

XX Claim 5; Page 97-98; 98pp; English.

XX The present sequence is human carbohydrate-associated protein (CARBAP)-2  
 CC cDNA. The CARBAP and the polynucleotide encoding it are useful for  
 CC diagnosing, treating and preventing cell proliferative disorders (e.g.  
 CC arteriosclerosis, cirrhosis, hepatitis, cancer), autoimmune/inflammatory  
 CC disorders (e.g. allergies, anaemia, asthma, osteoporosis, rheumatoid  
 CC arthritis, atopic dermatitis, glomerulonephritis and irritable bowel  
 CC syndrome, trauma, and bacterial, viral, parasitic, protozoal, fungal or  
 CC helminthic infections), reproductive disorders (e.g. infertility,  
 CC ovulatory defects and fibrocystic breast disease), and neurological  
 CC disorders (e.g. epilepsy, stroke, Alzheimer's disease, Huntington's  
 CC disease, dementia, Parkinson's disease, multiple sclerosis, mental  
 CC disorders including mood, anxiety and schizophrenic disorders, kuru,  
 CC bacterial and viral meningitis, Pick's disease and amyotrophic lateral  
 CC sclerosis). The polypeptide of the invention is also useful for screening  
 CC agonist, antagonist, a compound that specifically binds to it or  
 CC modulates its activity, for preparing polyclonal or monoclonal  
 CC antibodies, in a number of drug screening techniques, to analyse the  
 CC presence of a tissue or cell type and as element on a microarray.  
 CC The polynucleotide is useful for creating knockin humanised animals or  
 CC transgenic animals to model human diseases, in somatic or germline gene  
 CC therapy, to generate a transcript image of a tissue or cell type, for  
 CC detecting differences in the chromosomal location due to translocation,  
 CC inversion, among normal, carrier or affected individuals, and as

CC hybridisation probes for mapping naturally occurring genomic sequences.  
 XX  
 SQ Sequence 3239 BP; 659 A; 1005 C; 814 G; 761 T; 0 other;

Query Match 2.1%; Score 29; DB 24; Length 3239;  
 Best Local Similarity 100.0%; Pred. No. 0.071;  
 Matches 29; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1369 AAAAAAAAAAAAAAAAAAGCGGTC 1397  
 |||||  
 Db 38 AAAAAAAAAAAAAAAAAAGCGGTC 10

Search completed: June 29, 2003, 05:30:21  
 job time : 268 secs

















977 23 1.6 1013 3 US-09-248-335-71. Sequence 71, Appl  
978 23 1.6 1018 4 US-09-302-769-39 Sequence 39, Appl  
979 23 1.6 1020 4 US-09-425-578-5 Sequence 5, Appl  
980 23 1.6 1029 1 US-09-031-485-6 Sequence 6, Appl  
981 23 1.6 1029 1 US-09-031-485-8 Sequence 8, Appl  
982 23 1.6 1029 1 US-08-847-429A-6 Sequence 6, Appl  
983 23 1.6 1029 1 US-08-847-429A-8 Sequence 8, Appl  
984 23 1.6 1029 3 US-09-065-474-6 Sequence 6, Appl  
985 23 1.6 1029 3 US-09-065-474-8 Sequence 8, Appl  
986 23 1.6 1029 4 US-09-557-034-6 Sequence 8, Appl  
987 23 1.6 1029 4 US-09-557-034-8 Sequence 8, Appl  
988 23 1.6 1035 2 US-08-580-545B-3 Sequence 3, Appl  
989 23 1.6 1035 4 US-09-262-653A-3 Sequence 3, Appl  
990 23 1.6 1035 4 US-08-867-484A-1 Sequence 1, Appl  
991 23 1.6 1037 2 US-08-824-405-1 Sequence 1, Appl  
992 23 1.6 1046 4 US-09-068-140A-5 Sequence 5, Appl  
993 23 1.6 1058 4 US-09-452-239-11 Sequence 11, Appl  
994 23 1.6 1066 1 US-08-157-101A-4 Sequence 4, Appl  
995 23 1.6 1074 3 US-09-248-335-67 Sequence 67, Appl  
996 23 1.6 1075 4 US-08-400-006B-6 Sequence 6, Appl  
997 23 1.6 1078 6 5223394 Patent No. 5223394  
998 23 1.6 1082 4 US-09-247-155-181 Patent No. 5223394  
999 23 1.6 1092 4 US-09-227-357-35 Sequence 181, App  
1000 23 1.6 1100 3 US-09-248-335-53 Sequence 53, Appl

## ALIGNMENTS

RESULT 1  
US-08-473-981A-5  
Sequence 5, Application US/08473981A  
Patent No. 5629162  
GENERAL INFORMATION:  
APPLICANT: defougerolles, Antonin R  
TITLE OF INVENTION: METHODS OF IDENTIFYING AGENTS WHICH MODULATE  
TITLE OF INVENTION: ICAM-3 BINDING TO LFA-1 (AS AMENDED)  
NUMBER OF SEQUENCES: 14  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: STERNE, KESSLER, GOLDSTEIN & FOX, P.L.L.C.  
STREET: 1100 NEW YORK AVENUE, N.W. SUITE 600  
CITY: WASHINGTON  
STATE: D. C.  
COUNTRY: USA  
ZIP: 20005  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patentin Release #1.0, Version #1.30  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/473,981A  
FILING DATE: 07-JUN-1995  
CLASSIFICATION: 435  
ATTORNEY/AGENT INFORMATION:  
NAME: MILTONIG, ROBERT C  
REGISTRATION NUMBER: 34,395  
REFERENCE/DOCKET NUMBER: 1011.0560004  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (202) 371-2600  
TELEFAX: (202) 371-2540  
INFORMATION FOR SEQ ID NO: 5:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 1817 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: both  
TOPOLOGY: both  
MOLECULE TYPE: DNA (genomic)  
FEATURE:  
NAME/KEY: CDS  
LOCATION: 9..1649  
US-08-473-981A-5

Query Match 1.9%; Score 27; DB 1; Length 1817;  
Best Local Similarity 100.0%; Pred. No. 0.017;  
Matches 27; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1364 TCCCTAAAAAAAAAAAAAAAAAAAAA 1390  
|||||  
Db 1727 TCCCTAAAAAAAAAAAAAAAAAAAAA 1753

RESULT 2  
US-08-474-087-5  
Sequence 5, Application US/08474087  
Patent No. 5891841  
GENERAL INFORMATION:  
APPLICANT: de Fougerolles, Antonin R  
TITLE OF INVENTION: METHODS OF USING INTERCELLULAR ADHESION MOLECULE-  
NUMBER OF SEQUENCES: 14  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: STERNE, KESSLER, GOLDSTEIN & FOX, P.L.L.C.  
STREET: 1100 NEW YORK AVENUE, N.W. SUITE 600  
CITY: WASHINGTON  
STATE: D. C.  
COUNTRY: USA  
ZIP: 20005  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patentin Release #1.0, Version #1.30  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/474,087  
FILING DATE: 07-JUN-1995  
CLASSIFICATION: 514  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 08/038,990  
FILING DATE: 23-DEC-1992  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 07/712,879  
FILING DATE: 11-JUN-1991  
ATTORNEY/AGENT INFORMATION:  
NAME: MILTONIG, ROBERT C  
REGISTRATION NUMBER: 34,395  
REFERENCE/DOCKET NUMBER: 1011.0560003  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (202) 371-2600  
TELEFAX: (202) 371-2540  
INFORMATION FOR SEQ ID NO: 5:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 1817 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: both  
TOPOLOGY: both  
MOLECULE TYPE: DNA (genomic)  
FEATURE:  
NAME/KEY: CDS  
LOCATION: 9..1649  
US-08-474-087-5

Query Match 1.9%; Score 27; DB 2; Length 1817;  
Best Local Similarity 100.0%; Pred. No. 0.017;  
Matches 27; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1364 TCCCTAAAAAAAAAAAAAAAAAAAAA 1390  
|||||  
Db 1727 TCCCTAAAAAAAAAAAAAAAAAAAAA 1753

RESULT 3  
US-09-130-491-7  
Sequence 7, Application US/09130491  
Patent No. 6416974

```

; GENERAL INFORMATION:
; APPLICANT: Holtzman, Douglas A.
; APPLICANT: Goodearl, Andrew D. J.
; TITLE OF INVENTION: TANGO-71, TANGO-73, TANGO-74, TANGO-76, AND TANGO-83
; FILE REFERENCE: 09404/041001.
; CURRENT APPLICATION NUMBER: US/09/130,491
; EARLIER FILING DATE: 1998-08-07
; EARLIER APPLICATION NUMBER: US 60/058,108
; EARLIER FILING DATE: 1997-09-05
; EARLIER APPLICATION NUMBER: US 60/054,961
; EARLIER FILING DATE: 1997-08-06
; NUMBER OF SEQ ID NOS: 16
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 7
; LENGTH: 2114
; TYPE: DNA
; ORGANISM: Rattus rattus
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (3)...(1445)
; US-09-130-491-7

Query Match          1.9%; Score 27; DB 4; Length 2114;
Best Local Similarity 100.0%; Pred. No. 0.017;
Matches 27; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY      1369 AAAAAAAAAAAAAAAAAAGGCGG 1395
DB      2084 AAAAAAAAAAAAAAAAAAGGCGG 2110

RESULT 4
US-09-798-096-10/C
; Sequence 10, Application US/09798096
; Patent No. 639378
; GENERAL INFORMATION:
; APPLICANT: Donna T. Ward
; APPLICANT: Andrew T. Walt
; TITLE OF INVENTION: ANTISENSE MODULATION OF RECOL2 EXPRESSION
; FILE REFERENCE: RTS-0207
; CURRENT APPLICATION NUMBER: US/09/798,096
; CURRENT FILING DATE: 2001-03-01
; NUMBER OF SEQ ID NOS: 89
; SEQ ID NO 10
; LENGTH: 99500
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; US-09-798-096-10

Query Match          1.9%; Score 27; DB 4; Length 99500;
Best Local Similarity 100.0%; Pred. No. 0.014;
Matches 27; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY      1367 CTAATAAAAAAAAAAAAAAAAAAGGC 1393
DB      57302 CTAATAAAAAAAAAAAAAAAAAAGGC 57276

RESULT 5
US-09-122-400B-11
; Sequence 11, Application US/09122400B
; Patent No. 6245974
; GENERAL INFORMATION:
; APPLICANT: Michalowski, Susan
; APPLICANT: Spiker, Steven
; TITLE OF INVENTION: MATRIX ATTACHMENT REGIONS
; FILE REFERENCE: Michalowski and Spiker
; CURRENT APPLICATION NUMBER: US/09/122,400B
; CURRENT FILING DATE: 1998-07-24
; PRIOR APPLICATION NUMBER: 60/066,118
; PRIOR FILING DATE: 1997-08-06
; NUMBER OF SEQ ID NOS: 22
```

```

; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 11
; LENGTH: 899
; TYPE: DNA
; ORGANISM: Nicotiana tabacum
; US-09-122-400B-11

Query Match          1.9%; Score 26; DB 4; Length 899;
Best Local Similarity 100.0%; Pred. No. 0.047;
Matches 26; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY      1365 CCTTAAAAAAAAAAAAAAAAAAAAA 1390
DB      804 CCTTAAAAAAAAAAAAAAAAAAAAA 829

RESULT 6
US-09-257-179-28
; Sequence 28, Application US/09257179
; Patent No. 6410709
; GENERAL INFORMATION:
; APPLICANT: Ruben et al.
; TITLE OF INVENTION: 29 Human Secreted Proteins
; FILE REFERENCE: P2015Pl
; CURRENT APPLICATION NUMBER: US/09/257,179
; CURRENT FILING DATE: 1999-02-25
; EARLIER APPLICATION NUMBER: PCT/US98/17709
; EARLIER FILING DATE: 1998-08-27
; EARLIER APPLICATION NUMBER: 60/056,270
; EARLIER FILING DATE: 1997-08-29
; EARLIER APPLICATION NUMBER: 60/056,271
; EARLIER FILING DATE: 1997-08-29
; EARLIER APPLICATION NUMBER: 60/056,247
; EARLIER FILING DATE: 1997-08-29
; EARLIER APPLICATION NUMBER: 60/056,073
; EARLIER FILING DATE: 1997-08-29
; NUMBER OF SEQ ID NOS: 128
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 28
; LENGTH: 1327
; TYPE: DNA
; ORGANISM: Homo sapiens
; US-09-257-179-28

Query Match          1.9%; Score 26; DB 4; Length 1327;
Best Local Similarity 100.0%; Pred. No. 0.046;
Matches 26; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY      1370 AAAAAAAAAAAAAAAAAAGGCGG 1395
DB      1300 AAAAAAAAAAAAAAAAAAGGCGG 1325

RESULT 7
US-08-836-567-3
; Sequence 3, Application US/08836567
; Patent No. 6130367
; GENERAL INFORMATION:
; APPLICANT: Kossmann, Jens
; APPLICANT: Springer, Franziska
; APPLICANT: Abel, Gernot
; TITLE OF INVENTION: DNA MOLECULES THAT CODE FOR ENZYMES
; INVOLVED IN STARCH SYNTHESIS VECTORS BACTERIA TRANSGENIC
; TITLE OF INVENTION: PLANT CELLS AND PLANTS CONTAINING SAID MOLECULES
; NUMBER OF SEQUENCES: 17
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: FISH & NEAVE
; STREET: 1251 Avenue of the Americas
; CITY: New York
; STATE: New York
; COUNTRY: USA
; ZIP: 10020
; COMPUTER READABLE FORM:
```

MEDIUM TYPE: Floppy/disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patentin Release #1.0, Version #1.30  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/836,567  
FILING DATE: 24-JUL-1997  
CLASSIFICATION: 800  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: PCT/EP95/04415  
FILING DATE: 09-NOV-1995  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: IDE P 44 41 408.0  
FILING DATE: 10-NOV-1994  
ATTORNEY/AGENT INFORMATION:  
NAME: Haley Jr., James F.  
REGISTRATION NUMBER: 27,794  
REFERENCE/DOCKET NUMBER: Agrevo-4  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 212-596-9000  
TELEFAX: 212-596-9090  
INFORMATION FOR SEQ ID NO: 3:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 1758 base pairs  
TYPE: nucleotide  
STRANDEDNESS: unknown  
TOPOLOGY: linear  
MOLECULE TYPE: cDNA to mRNA  
HYPOTHETICAL: NO  
ANTI-SENSE: NO  
ORIGINAL SOURCE:  
ORGANISM: Solanum tuberosum  
STRAIN: cv. Bercolina  
TISSUE TYPE: tuber tissue  
IMMEDIATE SOURCE:  
LIBRARY: cDNA-library in pluescriptskit+  
FEATURE:  
NAME/KEY: CDS  
LOCATION: 1..1377  
OTHER INFORMATION: /function= "Polymerization of starch"  
OTHER INFORMATION: /product= "Starch synthase"  
US-08-836-567-3

Query Match 1.9%; Score 26; DB 3; Length 1758;  
Best Local Similarity 100.0%; Pred. No. 0.045;  
Matches 26; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1365 CCCTAAAAAAAAAAAAAAAAAAAA 1390  
DB 1719 CCTTAAAAAAAAAAAAAAAAAAAA 1744

RESULT 8  
US-08-287-001A-1  
Sequence 1, Application US/08287001A  
Patent No. 5622861  
GENERAL INFORMATION:  
APPLICANT: KAPLAN, GERARDO  
APPLICANT: FEINSTONE, STEPHEN M.  
TITLE OF INVENTION: HEPATITIS A VIRUS RECEPTOR AND METHODS  
TITLE OF INVENTION: OF USE  
NUMBER OF SEQUENCES: 5  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: NEEDLE & ROSENBERG, P.C.  
STREET: Suite 1200, The Candler Bldg, 127 Peachtree  
STREET: Street  
CITY: Atlanta  
STATE: Georgia  
COUNTRY: USA  
ZIP: 30303  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk

COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patentin Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/287,001A  
FILING DATE:  
CLASSIFICATION: 435  
ATTORNEY/AGENT INFORMATION:  
NAME: Spratt, Gwedolyn D.  
REGISTRATION NUMBER: 36,016  
REFERENCE/DOCKET NUMBER: 1414.621  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 404/688-0770  
TELEFAX: 404/688-9880  
INFORMATION FOR SEQ ID NO: 1:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 2093 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: double  
TOPOLOGY: linear  
MOLECULE TYPE: DNA  
FEATURE:  
NAME/KEY: CDS  
LOCATION: 196..1551  
US-08-287-001A-1

Query Match 1.9%; Score 26; DB 1; Length 2093;  
Best Local Similarity 100.0%; Pred. No. 0.045;  
Matches 26; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1365 CCCTAAAAAAAAAAAAAAAAAAAA 1390  
DB 2068 CCTTAAAAAAAAAAAAAAAAAAAA 2093

RESULT 9  
PCT-US95-09941-1  
Sequence 1, Application PC/TUS9509941  
GENERAL INFORMATION:  
APPLICANT:  
TITLE OF INVENTION: HEPATITIS A VIRUS RECEPTOR AND METHODS  
TITLE OF INVENTION: OF USE  
NUMBER OF SEQUENCES: 5  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: NEEDLE & ROSENBERG, P.C.  
STREET: Suite 1200, The Candler Bldg, 127 Peachtree  
STREET: Street  
CITY: Atlanta  
STATE: Georgia  
COUNTRY: USA  
ZIP: 30303  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patentin Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: PCT/US95/09941  
FILING DATE:  
CLASSIFICATION:  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: 08/287,001  
FILING DATE: 5 AUG 1994  
ATTORNEY/AGENT INFORMATION:  
NAME: Spratt, Gwedolyn D.  
REGISTRATION NUMBER: 36,016  
REFERENCE/DOCKET NUMBER: 1414.621  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 404/688-0770  
TELEFAX: 404/688-9880  
INFORMATION FOR SEQ ID NO: 1:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 2093 base pairs



TYPE: nucleic acid  
STRANDEDNESS: double  
TOPOLOGY: linear  
MOLECULE TYPE: DNA (genomic)  
FEATURE:  
NAME/KEY: CDS  
LOCATION: 196..1551  
PCT-US95-09941-1

Query Match 1.9%; Score 26; DB 5; Length 2093;  
Best Local Similarity 100.0%; Pred. No. 0.045;  
Matches 26; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1365 CCTAAAAAAAAAAAAAAAAAAAAA 1390  
Db 2068 CCTAAAAAAAAAAAAAAAAAAAAA 2093

RESULT 10  
US-08-969-630-3  
Sequence 3, Application US/08969630A  
Patent No. 5981248  
GENERAL INFORMATION:  
APPLICANT: Xu, Hua  
TITLE OF INVENTION: MAMMALIAN CELL DEATH PREVENTING KINASE, DPK  
FILE REFERENCE: A463  
CURRENT APPLICATION NUMBER: US/08/969,630A  
CURRENT FILING DATE: 1997-11-13  
NUMBER OF SEQ ID NOS: 7  
SOFTWARE: PatentIn Ver. 2.0  
SEQ ID NO 3  
LENGTH: 3891  
TYPE: DNA  
ORGANISM: mouse  
FEATURE:  
NAME/KEY: CDS  
LOCATION: (63)..(1565)  
US-08-969-630-3

Query Match 1.9%; Score 26; DB 2; Length 3891;  
Best Local Similarity 100.0%; Pred. No. 0.043;  
Matches 26; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1365 CCTAAAAAAAAAAAAAAAAAAAAA 1390  
Db 3861 CCTAAAAAAAAAAAAAAAAAAAAA 3886

RESULT 11  
US-08-469-260A-11/c  
Sequence 11, Application US/08469260A  
Patent No. 6451578  
GENERAL INFORMATION:  
APPLICANT: JOHN N. SIMONS  
APPLICANT: TAMI J. PILOT-MATIAS  
APPLICANT: GEORGE J. DAWSON  
APPLICANT: GEORGE G. SCHLAUDER  
APPLICANT: SURESH M. DESAI  
APPLICANT: THOMAS P. LEARY  
APPLICANT: ANTHONY SCOTT MIERHOFF  
APPLICANT: JAMES C. ERKER  
APPLICANT: SHERI L. BUIK  
APPLICANT: ISA K. MUSHAWAR  
TITLE OF INVENTION: NON-A, NON-B, NON-C, NON-D, NON-E HEPATITIS  
TITLE OF INVENTION: REAGENTS AND METHODS FOR THEIR USE  
NUMBER OF SEQUENCES: 716  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: ABBOTT LABORATORIES D377/AP6D  
STREET: 100 ABBOTT PARK ROAD  
CITY: ABBOTT PARK  
STATE: IL  
COUNTRY: USA  
ZIP: 60064-3500

COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: PatentIn Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/469,260A  
FILING DATE:  
CLASSIFICATION:  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US/08/424,550  
FILING DATE:  
ATTORNEY/AGENT INFORMATION:  
NAME: POREMBSKI, PRISCILLA E.  
REGISTRATION NUMBER: 33,207  
REFERENCE/DOCKET NUMBER: 5527.PC.01  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 708-937-6365  
TELEFAX: 708-938-2623  
INFORMATION FOR SEQ ID NO: 11:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 8912 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
MOLECULE TYPE: DNA (genomic)  
US-08-469-260A-11

Query Match 1.9%; Score 26; DB 4; Length 8912;  
Best Local Similarity 100.0%; Pred. No. 0.041;  
Matches 26; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1365 CCTAAAAAAAAAAAAAAAAAAAAA 1390  
Db 8911 CCTAAAAAAAAAAAAAAAAAAAAA 8886

RESULT 12  
US-08-639-857-32/c  
Sequence 32, Application US/08639857  
Patent No. 5955318  
GENERAL INFORMATION:  
APPLICANT: Simons, J. N.  
APPLICANT: Desai, S. M.  
APPLICANT: Mushawar, I. K.  
TITLE OF INVENTION: REAGENTS AND METHODS USEFUL FOR CONTROLLING THE  
TITLE OF INVENTION: TRANSLATION OF HEPATITIS GB PROTEINS  
NUMBER OF SEQUENCES: 32  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Abbott Laboratories  
STREET: 100 Abbott Park Rd  
CITY: Abbott Park  
STATE: IL  
COUNTRY: USA  
ZIP: 60064  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: PatentIn Release #1.0, Version #1.30  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/639,857  
FILING DATE:  
CLASSIFICATION: 435  
ATTORNEY/AGENT INFORMATION:  
NAME: Porembski, Priscilla E.  
REGISTRATION NUMBER: 33,207  
REFERENCE/DOCKET NUMBER: 5793.US.P1  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 708-937-0378  
TELEFAX: 708-938-2623  
INFORMATION FOR SEQ ID NO: 32:  
SEQUENCE CHARACTERISTICS:



LENGTH: 9143 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: double  
TOPOLOGY: linear  
MOLECULE TYPE: DNA (genomic)  
US-08-639-857-32

Query Match 1.9%; Score 26; DB 2; Length 9143;  
Best Local Similarity 100.0%; Pred. No. 0.041;  
Matches 26; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1365 CCTAAAAAAAAAAAAAAAAAAAAA 1390  
DB 9098 CCTAAAAAAAAAAAAAAAAAAAAA 9073

RESULT 13  
US-08-469-260A-390/C  
Sequence 390, Application US/08469260A  
Patent No. 6451578

GENERAL INFORMATION:  
APPLICANT: JOHN N. SIMONS  
APPLICANT: TAMI J. PILOT-MATIAS  
APPLICANT: GEORGE J. DAMSON  
APPLICANT: GEORGE G. SCHLAUDER  
APPLICANT: SURESH M. DESAI  
APPLICANT: THOMAS P. LEARY  
APPLICANT: ANTHONY SCOTT MUEHROFF  
APPLICANT: JAMES C. ERKER  
APPLICANT: SHERI L. BUTIK  
APPLICANT: ISA K. MUSHAMMAR  
TITLE OF INVENTION: NON-A, NON-B, NON-C, NON-D, NON-E HEPATITIS  
TITLE OF INVENTION: REAGENTS AND METHODS FOR THEIR USE  
NUMBER OF SEQUENCES: 716  
CORRESPONDENCE ADDRESSES:  
ADDRESSEE: ABBOTT LABORATORIES D377/AP6D  
STREET: 100 ABBOTT PARK ROAD  
CITY: ABBOTT PARK  
STATE: IL  
COUNTRY: USA  
ZIP: 60064-3500  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patentin Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/469,260A  
FILING DATE:  
CLASSIFICATION:  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US/08/424,550  
FILING DATE:  
ATTORNEY/AGENT INFORMATION:  
NAME: FOREMSKI, PRISCILLA E.  
REGISTRATION NUMBER: 33,207  
REFERENCE/DOCKET NUMBER: 5527.PC.01  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 708-937-6365  
TELEFAX: 708-938-2623  
INFORMATION FOR SEQ ID NO: 390:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 9143 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: double  
TOPOLOGY: linear  
MOLECULE TYPE: DNA (genomic)  
US-08-469-260A-390.

Query Match 1.9%; Score 26; DB 4; Length 9143;  
Best Local Similarity 100.0%; Pred. No. 0.041;  
Matches 26; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1365 CCTAAAAAAAAAAAAAAAAAAAAA 1390  
DB 9098 CCTAAAAAAAAAAAAAAAAAAAAA 9073

RESULT 14  
US-08-469-260A-393/C  
Sequence 393, Application US/08469260A  
Patent No. 6451578

GENERAL INFORMATION:  
APPLICANT: JOHN N. SIMONS  
APPLICANT: TAMI J. PILOT-MATIAS  
APPLICANT: GEORGE J. DAMSON  
APPLICANT: GEORGE G. SCHLAUDER  
APPLICANT: SURESH M. DESAI  
APPLICANT: THOMAS P. LEARY  
APPLICANT: ANTHONY SCOTT MUEHROFF  
APPLICANT: JAMES C. ERKER  
APPLICANT: SHERI L. BUTIK  
APPLICANT: ISA K. MUSHAMMAR  
TITLE OF INVENTION: NON-A, NON-B, NON-C, NON-D, NON-E HEPATITIS  
TITLE OF INVENTION: REAGENTS AND METHODS FOR THEIR USE  
NUMBER OF SEQUENCES: 716  
CORRESPONDENCE ADDRESSES:  
ADDRESSEE: ABBOTT LABORATORIES D377/AP6D  
STREET: 100 ABBOTT PARK ROAD  
CITY: ABBOTT PARK  
STATE: IL  
COUNTRY: USA  
ZIP: 60064-3500  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patentin Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/469,260A  
FILING DATE:  
CLASSIFICATION:  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US/08/424,550  
FILING DATE:  
ATTORNEY/AGENT INFORMATION:  
NAME: FOREMSKI, PRISCILLA E.  
REGISTRATION NUMBER: 33,207  
REFERENCE/DOCKET NUMBER: 5527.PC.01  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 708-937-6365  
TELEFAX: 708-938-2623  
INFORMATION FOR SEQ ID NO: 393:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 9143 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: double  
TOPOLOGY: linear  
MOLECULE TYPE: DNA (genomic)  
FEATURE:  
NAME/KEY: 5'UTR  
LOCATION: 1..445  
FEATURE:  
NAME/KEY: CDS  
LOCATION: 446..9037  
FEATURE:  
NAME/KEY: 3'UTR  
LOCATION: 9038..9143  
US-08-469-260A-393

Query Match 1.9%; Score 26; DB 4; Length 9143;  
Best Local Similarity 100.0%; Pred. No. 0.041;  
Matches 26; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1365 CCTAAAAAAAAAAAAAAAAAAAAA 1390

Db 9098 CCTAAAAAAAAAAAAAAAAAAAAA 9073

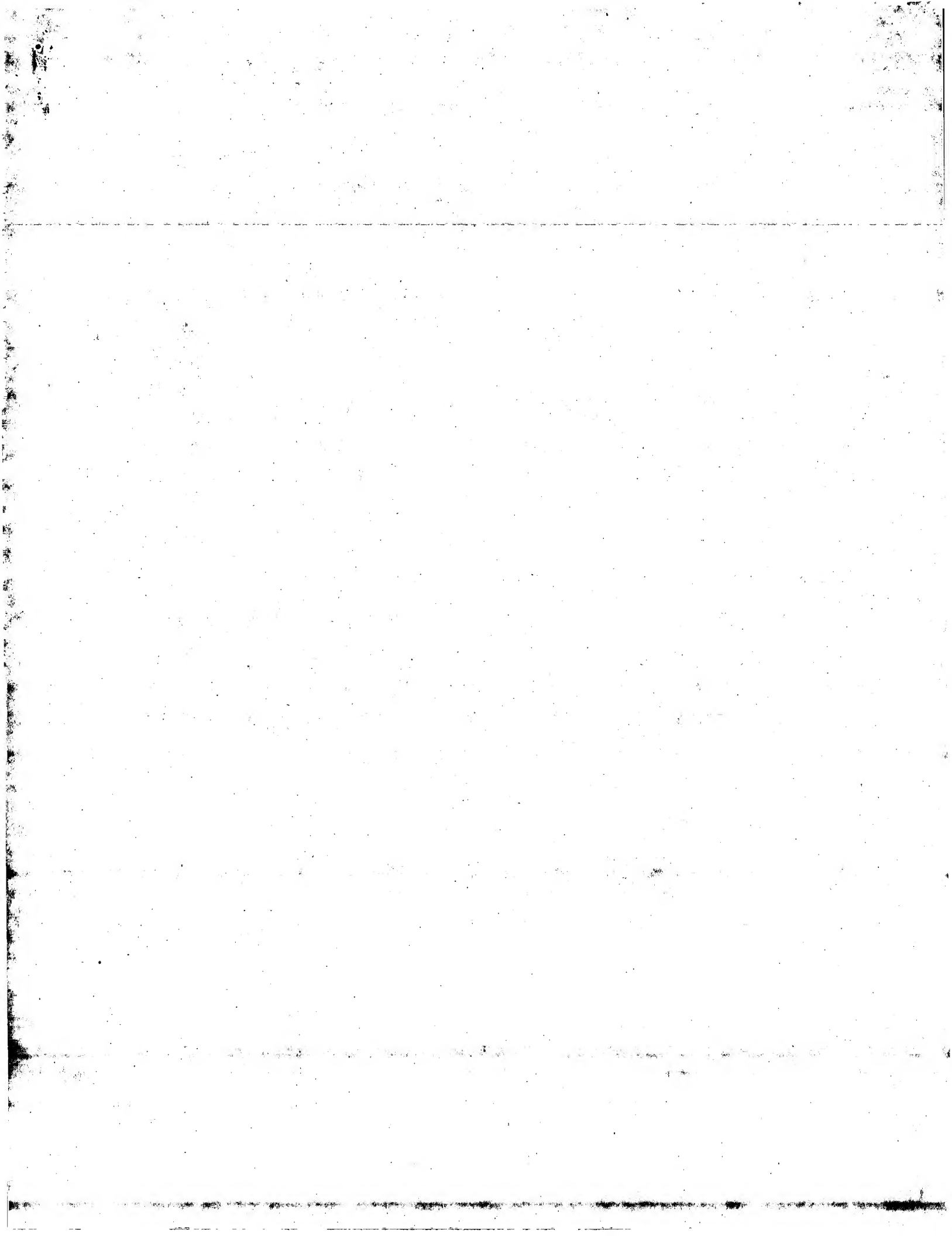
RESULT 15

US-09-821-736-3  
 ; Sequence 3, Application US/09821736  
 ; Patent No. 6326182  
 ; GENERAL INFORMATION:  
 ; APPLICANT: WEBSTER, Marion et al  
 ; TITLE OF INVENTION: ISOLATED HUMAN LIPASE PROTEINS, NUCLEIC  
 ; TITLE OF INVENTION: ACID MOLECULES ENCODING HUMAN LIPASE PROTEINS, AND USES  
 ; TITLE OF INVENTION: THEREOF  
 ; FILE REFERENCE: CL001216  
 ; CURRENT APPLICATION NUMBER: US/09/821,736  
 ; CURRENT FILING DATE: 2001-03-30  
 ; NUMBER OF SEQ ID NOS: 5  
 ; SOFTWARE: FastSeq for Windows Version 4.0  
 ; SEQ ID NO 3  
 ; LENGTH: 14753  
 ; TYPE: DNA  
 ; ORGANISM: Human  
 ; FEATURE:  
 ; NAME/KEY: misc\_feature  
 ; LOCATION: (1)..(14753)  
 ; OTHER INFORMATION: n = A,T,C or G  
 US-09-821-736-3

Query Match 1.9%; Score 26; DB 4; Length 14753;  
 Best Local Similarity 100.0%; Pred.No.0.04;  
 Matches 26; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1366 CCTAAAAAAAAAAAAAAAAAAAAAAG 1391  
 Db 2131 CCTAAAAAAAAAAAAAAAAAAAAAAG 2156

Search completed: June 29, 2003, 06:37:45  
 Job time : 75 secs



Mon Jun 30 09:37:15 2003

GenCore version 5.1.6  
Copyright (c) 1993 - 2003 Compen Ltd.

OM nucleic - nucleic search, using sw model

Run on: June 29, 2003, 05:25:43 ; Search time 166 Seconds  
(without alignments)  
12492.616 Million cell updates/sec

Title: US-09-989-919-15

Perfect score: 1397

Sequence: 1 ggtcgtcactctacgga.....aaaaaaaaaagcggtc 1397

Scoring table: OLIGO NUC  
Gapop\_60.0, Gapext 60.0

Searched: 1055720 seqs, 742224136 residues

Word size: 0  
Total number of hits satisfying chosen parameters: 2111440Minimum DB seq length: 0  
Maximum DB seq length: 200000000

Post-processing: Listing first 1000 summaries

Database:

```
1: /cgn2_6/ptodata/1/pubpna/US07_PUBCOMB.seq:*
2: /cgn2_6/ptodata/1/pubpna/PCT_NEW_PUB.seq:*
3: /cgn2_6/ptodata/1/pubpna/US06_NEW_PUB.seq:*
4: /cgn2_6/ptodata/1/pubpna/US07_NEW_PUB.seq:*
5: /cgn2_6/ptodata/1/pubpna/US07_PUBCOMB.seq:*
6: /cgn2_6/ptodata/1/pubpna/PCTUS_PUBCOMB.seq:*
7: /cgn2_6/ptodata/1/pubpna/US08_NEW_PUB.seq:*
8: /cgn2_6/ptodata/1/pubpna/US08_PUBCOMB.seq:*
9: /cgn2_6/ptodata/1/pubpna/US09_NEW_PUB.seq:*
10: /cgn2_6/ptodata/1/pubpna/US09_PUBCOMB.seq:*
11: /cgn2_6/ptodata/1/pubpna/US10_NEW_PUB.seq:*
12: /cgn2_6/ptodata/1/pubpna/US10_PUBCOMB.seq:*
13: /cgn2_6/ptodata/1/pubpna/US60_NEW_PUB.seq:*
14: /cgn2_6/ptodata/1/pubpna/US60_PUBCOMB.seq:*
```

Pred. No. is the number of results predicted by chance to have a  
score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	1397	100.0	4370	9	US-09-989-919-15
2	442	31.6	470	9	US-09-989-919-14
3	322	23.0	427	10	US-09-989-919-14
4	97	6.9	493	9	US-09-989-919-14
5	30	2.1	386	10	US-09-989-919-14
6	30	2.1	421	10	US-09-989-919-14
7	30	2.1	777	10	US-09-989-919-14
8	30	2.1	2457	9	US-10-176-482-569
9	30	2.1	2457	9	US-10-176-482-569
10	30	2.1	2457	9	US-10-176-482-569
11	30	2.1	2457	9	US-10-176-482-569
12	30	2.1	2457	9	US-10-176-482-569
13	30	2.1	2457	9	US-10-176-482-569
14	30	2.1	2457	9	US-10-176-482-569
15	30	2.1	2457	9	US-10-176-482-569
16	30	2.1	2457	9	US-10-176-482-569
17	30	2.1	2457	9	US-10-176-482-569
18	30	2.1	2457	9	US-10-176-482-569
19	30	2.1	2457	9	US-10-176-482-569















969	25	1.8	1992	9	US-10-036-542-18	Sequence 18, Appl
C 970	25	1.8	2000	9	US-09-938-842A-3635	Sequence 3635, Ap
971	25	1.8	2010	10	US-09-809-545A-52	Sequence 52, Appl
972	25	1.8	2014	9	US-10-106-698-390	Sequence 390, Ap
973	25	1.8	2085	9	US-10-106-698-2105	Sequence 2105, Ap
974	25	1.8	2098	9	US-10-163-866-11	Sequence 11, Appl
C 975	25	1.8	2122	9	US-10-001-873-2	Sequence 2, Appl
976	25	1.8	2124	9	US-10-106-698-1647	Sequence 1647, Ap
977	25	1.8	2209	9	US-10-198-846-13927	Sequence 13927, A
978	25	1.8	2223	9	US-10-125-540-110	Sequence 110, Appl
979	25	1.8	2223	10	US-09-764-870-110	Sequence 110, Appl
980	25	1.8	2229	10	US-09-925-301-160	Sequence 160, Appl
C 981	25	1.8	2273	10	US-09-925-301-257	Sequence 257, Appl
C 982	25	1.8	2290	10	US-09-822-849A-152	Sequence 152, Appl
C 983	25	1.8	2295	9	US-09-764-891-7825	Sequence 7825, Ap
984	25	1.8	2340	9	US-10-106-698-1073	Sequence 1073, Ap
985	25	1.8	2342	10	US-09-925-297-311	Sequence 311, Appl
986	25	1.8	2359	9	US-10-091-438-44	Sequence 44, Appl
987	25	1.8	2370	10	US-09-764-853-144	Sequence 144, Appl
988	25	1.8	2391	9	US-09-822-846-103	Sequence 103, Appl
989	25	1.8	2402	9	US-10-194-125-1	Sequence 1, Appl
990	25	1.8	2413	9	US-10-205-823-438	Sequence 438, Appl
991	25	1.8	2438	9	US-10-102-806-13	Sequence 13, Appl
992	25	1.8	2557	10	US-09-925-300-471	Sequence 471, Appl
993	25	1.8	2574	9	US-09-978-295A-101	Sequence 101, Appl
994	25	1.8	2574	9	US-09-978-697-101	Sequence 101, Appl
995	25	1.8	2574	9	US-09-978-192A-101	Sequence 101, Appl
996	25	1.8	2574	9	US-09-999-832A-101	Sequence 101, Appl
997	25	1.8	2574	9	US-09-978-189-101	Sequence 101, Appl
998	25	1.8	2574	9	US-09-978-608A-101	Sequence 101, Appl
999	25	1.8	2574	9	US-09-978-191A-101	Sequence 101, Appl
1000	25	1.8	2574	9	US-09-978-403A-101	Sequence 101, Appl

## ALIGNMENTS

RESULT 1  
 US-09-989-919-15  
 Sequence 15, Application US/09989919  
 Patent No. US2002016434A1  
 GENERAL INFORMATION:  
 APPLICANT: Macina, Roberto  
 APPLICANT: Recipon, Hervé  
 APPLICANT: Pluta, Jason  
 APPLICANT: Ghosh, Malaivika  
 APPLICANT: Sun, Yongming  
 APPLICANT: Liu, Chenghua  
 TITLE OF INVENTION: Compositions and Methods Relating to Colon Specific Genes and Pro  
 FILE REFERENCE: DEX-0289  
 CURRENT APPLICATION NUMBER: US/09/989,919  
 CURRENT FILING DATE: 2001-11-21  
 PRIOR APPLICATION NUMBER: 60/2252,505  
 PRIOR FILING DATE: 2000-11-22  
 NUMBER OF SEQ ID NOS: 124  
 SOFTWARE: PatentIn version 3.1  
 SEQ ID NO 15  
 LENGTH: 1397  
 TYPE: DNA  
 ORGANISM: Homo sapien  
 US-09-989-919-15

Query Match 100.0%; Score 1397; DB 9; Length 1397;  
 Best Local Similarity 100.0%; Pred. No. 0;  
 Matches 1397; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Db 1 CGATGCTGACCTGTACCGGAGCGGAGTATCTGAGAACTCCAGCGAGACAGACATAC 60  
 1 GGTGTGACCTCTGACCGGAGCGGAGTATCTGAGAACTCCAGCGAGACAGACATAC 60  
 61 CGAGTACAGTGTATCCAGACAGACATATCCCGAGAGAGACTACCGCTGTGCTATC 120  
 61 CGAGTACAGTGTATCCAGACAGACATATCCCGAGAGAGACTACCGCTGTGCTATC 120

121 CTACACACAGGGAGCTGCCCTCTTCAAGTGTCAACCTGGCTGAGCTGTGATGCTG 180  
 121 CTACACACAGGGAGCTGCCCTCTTCAAGTGTCAACCTGGCTGAGCTGTGATGCTG 180  
 121 CTACACACAGGGAGCTGCCCTCTTCAAGTGTCAACCTGGCTGAGCTGTGATGCTG 180  
 181 TGAGAGCCATCCAGAGTGTGGGCTTGTGTGTCACCAACCAACCTGAGAGGTGA 240  
 181 TGAGAGCCATCCAGAGTGTGGGCTTGTGTGTCACCAACCAACCTGAGAGGTGA 240  
 241 GCGAGTGGAGAGAGCTTCCAGAGGAGATGGCAGACCTTCTGAGAGTGTATGATG 300  
 241 GCGAGTGGAGAGAGCTTCCAGAGGAGATGGCAGACCTTCTGAGAGTGTATGATG 300  
 241 GCGAGTGGAGAGAGCTTCCAGAGGAGATGGCAGACCTTCTGAGAGTGTATGATG 300  
 301 TGATCCCTCCATCGGAAGTCAGAGGGGGTCTGAGTGTATGAGAGAGTATAGTGTCT 360  
 301 TGATCCCTCCATCGGAAGTCAGAGGGGGTCTGAGTGTATGAGAGAGTATAGTGTCT 360  
 301 TGATCCCTCCATCGGAAGTCAGAGGGGGTCTGAGTGTATGAGAGAGTATAGTGTCT 360  
 361 TCAGGAGAGTCAATTTAGGAGAGATGTCTTGTCTCCAGAAAGAGAAATCCAGCCTG 420  
 361 TCAGGAGAGTCAATTTAGGAGAGATGTCTTGTCTCCAGAAAGAGAAATCCAGCCTG 420  
 421 TTAACCTTCACTCTGCCCCCAGGTGGCAGCTGTCTTTTAAAGCTGATGAGACC 480  
 421 TTAACCTTCACTCTGCCCCCAGGTGGCAGCTGTCTTTTAAAGCTGATGAGACC 480  
 421 TTAACCTTCACTCTGCCCCCAGGTGGCAGCTGTCTTTTAAAGCTGATGAGACC 480  
 481 AGTGTGCTGATCCCAAGACACATATGTGAAGGCTGTGACTATCTGAG 540  
 481 AGTGTGCTGATCCCAAGACACATATGTGAAGGCTGTGACTATCTGAG 540  
 481 AGTGTGCTGATCCCAAGACACATATGTGAAGGCTGTGACTATCTGAG 540  
 541 GGGTGGCTGACAGCTATCTCTCAAGCTGTGGCTTGTGCTGTGAGAGAGTGAATT 600  
 541 GGGTGGCTGACAGCTATCTCTCAAGCTGTGGCTTGTGCTGTGAGAGAGTGAATT 600  
 541 GGGTGGCTGACAGCTATCTCTCAAGCTGTGGCTTGTGCTGTGAGAGAGTGAATT 600  
 601 GACATGAGACAGTCAATGTACCTGGGAGACCCCTGAGACAAAGCTTAATCCAGACA 660  
 601 GACATGAGACAGTCAATGTACCTGGGAGACCCCTGAGACAAAGCTTAATCCAGACA 660  
 601 GACATGAGACAGTCAATGTACCTGGGAGACCCCTGAGACAAAGCTTAATCCAGACA 660  
 661 GACATGAGACAGTCAATGTACCTGGGAGACCCCTGAGACAAAGCTTAATCCAGACA 720  
 661 GACATGAGACAGTCAATGTACCTGGGAGACCCCTGAGACAAAGCTTAATCCAGACA 720  
 721 CCTAGTATGAGACCTGTGGCTCTTCAAGTCAAGGATCATGGGGGTATATGCTCTTCCA 780  
 721 CCTAGTATGAGACCTGTGGCTCTTCAAGTCAAGGATCATGGGGGTATATGCTCTTCCA 780  
 721 CCTAGTATGAGACCTGTGGCTCTTCAAGTCAAGGATCATGGGGGTATATGCTCTTCCA 780  
 781 ACCGTGTGGGCTGTAGCAAGCTCAGGCTAGTCTCCCACTGGGGGTGTGCTCTTCCA 840  
 781 ACCGTGTGGGCTGTAGCAAGCTCAGGCTAGTCTCCCACTGGGGGTGTGCTCTTCCA 840  
 781 ACCGTGTGGGCTGTAGCAAGCTCAGGCTAGTCTCCCACTGGGGGTGTGCTCTTCCA 840  
 841 GGGAGCGTTCGTTGGGAGCCCATCTGCTGTTCAATGTGTGATGTGATGCTTAAC 900  
 841 GGGAGCGTTCGTTGGGAGCCCATCTGCTGTTCAATGTGTGATGTGATGCTTAAC 900  
 901 CCTGTGCTGTGCTGTGCAATGCAACAGCAGAGCGGTGGGGGTGTGGGGGACATGCA 960  
 901 CCTGTGCTGTGCTGTGCAATGCAACAGCAGAGCGGTGGGGGTGTGGGGGACATGCA 960  
 901 CCTGTGCTGTGCTGTGCAATGCAACAGCAGAGCGGTGGGGGTGTGGGGGACATGCA 960  
 961 TCGTGTGAGTGTCTCTCAGCTTAGTCTGTGAGACAGAGACTTGGCGGGAGTGTCCAGGA 1020  
 961 TCGTGTGAGTGTCTCTCAGCTTAGTCTGTGAGACAGAGACTTGGCGGGAGTGTCCAGGA 1020  
 961 TCGTGTGAGTGTCTCTCAGCTTAGTCTGTGAGACAGAGACTTGGCGGGAGTGTCCAGGA 1020  
 1021 TGTGGGTATTTCTGATCTGGGAGGCTATCTGTACCTCCGACAGGGGACATCTCCAG 1080  
 1021 TGTGGGTATTTCTGATCTGGGAGGCTATCTGTACCTCCGACAGGGGACATCTCCAG 1080  
 1021 TGTGGGTATTTCTGATCTGGGAGGCTATCTGTACCTCCGACAGGGGACATCTCCAG 1080  
 1081 GCGAGCCAGGGGTGACAGGGGACAGAGTGTGACACTGAGATGAGCCAGATGAGGCTGAG 1140  
 1081 GCGAGCCAGGGGTGACAGGGGACAGAGTGTGACACTGAGATGAGGCTGAG 1140  
 1081 GCGAGCCAGGGGTGACAGGGGACAGAGTGTGACACTGAGATGAGGCTGAG 1140  
 1141 GAGAGAGTGTGTTTGTGAGCAGAGACTGTGGGGGTGGGGGTGGGGGCTTTTGTGCTT 1200  
 1141 GAGAGAGTGTGTTTGTGAGCAGAGACTGTGGGGGTGGGGGTGGGGGCTTTTGTGCTT 1200

QY 1201 CATTGCTTCATGAAAGCTCAAGACCAAAACAGGCTTTCCCTCTCTGAGT 1260  
DB 1201 CATTGCTTCATGAAAGCTCAAGACCAAAACAGGCTTTCCCTCTCTGAGT 1260  
QY 1261 TTGAATATCAGATCTTTGTACTCTTGTGTTAAATGTTATTTTGTAAAAAT 1320  
DB 1261 TTGAATATCAGATCTTTGTACTCTTGTGTTAAATGTTATTTTGTAAAAAT 1320  
QY 1321 AAAATAAATTAATTAATAAATGATGTTTACAGAGAACTCTTCCCTAAAAAAA 1380  
DB 1321 AAAATAAATTAATTAATAAATGATGTTTACAGAGAACTCTTCCCTAAAAAAA 1380  
QY 1381 AAAAAAAAAAGCGGTC 1397  
DB 1381 AAAAAAAAAAGCGGTC 1397

RESULT 2  
US-09-989-919-14  
Sequence 14, Application US/09989919  
Patent No. US20020164344A1  
GENERAL INFORMATION:  
APPLICANT: Macina, Roberto  
APPLICANT: Recipon, Hervé  
APPLICANT: Pluta, Jason  
APPLICANT: Ghosh, Malavika  
APPLICANT: Sun, Yongming  
APPLICANT: Liu, Chenghua  
TITLE OF INVENTION: Compositions and Methods Relating to Colon Specific Genes and Pro  
FILE REFERENCE: DEX-0289  
CURRENT APPLICATION NUMBER: US/09/989,919  
CURRENT FILING DATE: 2001-11-21  
PRIOR APPLICATION NUMBER: 60/252,505  
PRIOR FILING DATE: 2000-11-22  
NUMBER OF SEQ ID NOS: 124  
SOFTWARE: PatentIn version 3.1  
SEQ ID NO 14  
LENGTH: 470  
TYPE: DNA  
ORGANISM: Homo sapien  
US-09-989-919-14

Query Match 31.6%; Score 442; DB 9; Length 470;  
Best Local Similarity 100.0%; Pred. No. 6,8e-211;  
Matches 442; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 572 GCTGGGCTTCCCTGTGAGGAGTGAATGCTGCTGAGCAGACATGCTGCTGAGGAG 631  
DB 7 GCTGGGCTTCCCTGTGAGGAGTGAATGCTGCTGAGCAGACATGCTGCTGAGGAG 66  
QY 632 CCTGAGACAAAGCTAACATCCAGACAGAGATGTGACAGAGCAAACTGCAATAA 691  
DB 67 CCTGAGACAAAGCTAACATCCAGACAGAGATGTGACAGAGCAAACTGCAATAA 126  
QY 692 TGGCAATGTTAAATGTGAGTTTACAGGCTTACGATGAGGATGCTGCTCTTCTGCA 751  
DB 127 TGGCAATGTTAAATGTGAGTTTACAGGCTTACGATGAGGATGCTGCTCTTCTGCA 186  
QY 752 GGAATCATGGGGATGATGCTGCTCTTCAACCTGTGGGCTGTAAAGAACTGAGCTAG 811  
DB 187 GGAATCATGGGGATGATGCTGCTCTTCAACCTGTGGGCTGTAAAGAACTGAGCTAG 246  
QY 812 TCTCCCACTGGGGGCTGTGCTCCCTGCTGAGGAGGCTGCTGGGAGCCCACTGCTG 871  
DB 247 TCTCCCACTGGGGGCTGTGCTCCCTGCTGAGGAGGCTGCTGGGAGCCCACTGCTG 306  
QY 872 GTTCAATAGTGTGAGAAATGTTAAAGCCCTGCTGCTGCTGCTGCTGCTGCTGCTG 931  
DB 307 GTTCAATAGTGTGAGAAATGTTAAAGCCCTGCTGCTGCTGCTGCTGCTGCTGCTG 366  
QY 932 GGGGTGGGGGCTGCTGGGGGCAATCCAGCTGAGAGTCTTCTAGCTTAGTCTGGA 991  
DB 367 GGGGTGGGGGCTGCTGGGGGCAATCCAGCTGAGAGTCTTCTAGCTTAGTCTGGA 426

QY 992 CAGAGACTTGGCGGGGATGC 1013  
DB 427 CAGAGACTTGGCGGGGATGC 448

RESULT 3  
US-09-880-107-1138/c  
Sequence 1138, Application US/09880107  
Patent No. US20020142981A1  
GENERAL INFORMATION:  
APPLICANT: Horne, Darci T.  
APPLICANT: Vockley, Joseph G.  
APPLICANT: Scherf, Uwe  
APPLICANT: Gene Logic, Inc.  
TITLE OF INVENTION: Gene Expression Profiles in Liver Cancer  
FILE REFERENCE: 44921-5028-WO  
CURRENT APPLICATION NUMBER: US/09/880,107  
CURRENT FILING DATE: 2001-06-14  
PRIOR APPLICATION NUMBER: US 60/211,379  
PRIOR FILING DATE: 2000-06-14  
PRIOR APPLICATION NUMBER: US 60/237,054  
PRIOR FILING DATE: 2000-10-02  
NUMBER OF SEQ ID NOS: 3950  
SOFTWARE: PatentIn Ver. 2.1  
SEQ ID NO 1138  
LENGTH: 427  
TYPE: DNA  
ORGANISM: Homo sapiens  
FEATURE:  
OTHER INFORMATION: Genbank Accession No. US20020142981A1 AA451877  
US-09-880-107-1138

Query Match 23.0%; Score 322; DB 10; Length 427;  
Best Local Similarity 99.5%; Pred. No. 7.1e-151;  
Matches 422; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 936 GTGGGGCTGCTGGGGGCAATCCATCCGAGAGTGTCTCTGAGCTTAGTCTGACAG 995  
DB 427 GTGGGGCTGCTGGGGGCAATCCATCCGAGAGTGTCTCTGAGCTTAGTCTGACAG 368  
QY 996 AGACTTGGCGGGGATGCTCCAGATGTGGTGAATTTCTTACCTGGGAGGCTATCTG 1055  
DB 367 AGACTTGGCGGGGATGCTCCAGATGTGGTGAATTTCTTACCTGGGAGGCTATCTG 308  
QY 1056 ACCCTCCGACAGGGGACATCCAGAGCCAGCCAGGGGCTCAGGGGAGAGGTGCAACT 1115  
DB 307 ACCCTCCGACAGGGGACATCCAGAGCCAGCCAGGGGCTCAGGGGAGAGGTGCAACT 248  
QY 1116 CAGCATGAGCCAAAGCTGGGGGTCAAGGAGCAGGTGTGTTGAGCCAGGACCTGGGGCGG 1175  
DB 247 CAGCATGAGCCAAAGCTGGGGGTCAAGGAGCAGGTGTGTTGAGCCAGGACCTGGGGCGG 188  
QY 1176 GGGTGGGGCGGGGCTTCTGCTCATTTGCTTCAATGAAGCTCAAGACAGCAAA 1235  
DB 187 GGGTGGGGCGGGGCTTCTGCTCATTTGCTTCAATGAAGCTCAAGACAGCAAA 128  
QY 1236 ACCGAGCTTCCCTCTCTGAGATTGAATTCAGAACTTTTGTACTTCTTGTGTGT 1295  
DB 127 ACCGAGCTTCCCTCTCTGAGATTGAATTCAGAACTTTTGTACTTCTTGTGTGT 68  
QY 1296 TAAATGTTATTTTGTAAAAATTAATAAATTAATTAATTAATTAATTAATTAATTAAT 1355  
DB 67 TAAATGTTATTTTGTAAAAATTAATAAATTAATTAATTAATTAATTAATTAATTAAT 8  
QY 1356 CAAA 1359  
DB 7 CAAA 4

RESULT 4  
US-09-918-995-32213  
Sequence 32213, Application US/09918995

Publication No. US20030073623A1  
GENERAL INFORMATION:  
APPLICANT: Hyseq, Inc.  
TITLE OF INVENTION: NOVEL NUCLEIC ACID SEQUENCES OBTAINED  
FROM VARIOUS CDNA LIBRARIES  
FILE REFERENCE: 20411-756  
CURRENT APPLICATION NUMBER: US/09/918,995  
CURRENT FILING DATE: 2001-07-30  
PRIOR APPLICATION NUMBER: US/09/235,076  
PRIOR FILING DATE: 1999-01-20  
NUMBER OF SEQ ID NOS: 38054  
SOFTWARE: FastSeq for Windows Version 3.0  
SEQ ID NO 32213  
LENGTH: 493  
TYPE: DNA  
ORGANISM: Homo sapiens  
FEATURE:  
NAME/KEY: misc\_feature  
LOCATION: (1)...(493)  
OTHER INFORMATION: n = A,T,C or G  
US-09-918-995-32213

Query Match 6.9%; Score 97; DB 9; Length 493;  
Best Local Similarity 100.0%; Pred. No. 2,4e-38;  
Matches 97; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 GGTCGTGACCTGTACCGAGCGGAGCATCTGCAGAACTCCACGGCAGCAGCAGTAC 60  
DB 397 GGTCGTGACCTGTACCGAGCGGAGCATCTGCAGAACTCCACGGCAGCAGCAGTAC 456  
DB 457 CGAGTACCAGTGTATCCCGACAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAG 493

RESULT 5  
US-09-770-791-70/c  
Sequence 70, Application US/0970791  
Patent No. US2002062014A1

GENERAL INFORMATION:  
APPLICANT: Gorlach, Jorn  
APPLICANT: An, Yong-Qiang  
APPLICANT: Hamilton, Carol M.  
APPLICANT: Price, Jennifer L.  
APPLICANT: Raines, Tracy M.  
APPLICANT: Yu, Yang  
APPLICANT: Rameaka, Joshua G.  
APPLICANT: Page, Amy  
APPLICANT: Matthew, Abraham V.  
APPLICANT: Ledford, Brooke L.  
APPLICANT: Moesner, Jeffrey P.  
APPLICANT: Haas, William David  
APPLICANT: Garcia, Carlos A.  
APPLICANT: Krickler, Maja  
APPLICANT: Slader, Ted  
APPLICANT: Davis, Keith R.  
APPLICANT: Allen, Keith  
APPLICANT: Hoffman, Neil  
APPLICANT: Hubman, Patrick

TITLE OF INVENTION: Expressed Sequences of Arabidopsis  
TITLE OF INVENTION: Chailana  
FILE REFERENCE: 2029 (PARA-018PRV)  
CURRENT APPLICATION NUMBER: US/09/770,791  
CURRENT FILING DATE: 2001-01-26  
PRIOR APPLICATION NUMBER: 60/178,480  
PRIOR FILING DATE: 2000-01-27  
NUMBER OF SEQ ID NOS: 999  
SOFTWARE: FastSeq for Windows Version 4.0  
SEQ ID NO 70  
LENGTH: 386  
TYPE: DNA  
ORGANISM: Arabidopsis thaliana  
US-09-770-791-70

Query Match 2.1%; Score 30; DB 10; Length 386;  
Best Local Similarity 100.0%; Pred. No. 8e-05;  
Matches 30; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1361 TCTTCCCTTAAAAAAAAAAAAAAAAAAAAA 1390  
DB 31 TCTTCCCTTAAAAAAAAAAAAAAAAAAAAA 2

RESULT 6  
US-09-867-701-3641/c  
Sequence 3641, Application US/09867701  
Patent No. US2002032237A1  
GENERAL INFORMATION:  
APPLICANT: Aglate, Paul A.  
APPLICANT: Jones, Robert  
APPLICANT: Harlocker, Susan L.  
TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE THERAPY  
OF OVARIAN CANCER  
FILE REFERENCE: 210121.497  
CURRENT APPLICATION NUMBER: US/09/867,701  
CURRENT FILING DATE: 2001-05-29  
NUMBER OF SEQ ID NOS: 10912  
SOFTWARE: FastSeq for Windows Version 4.0  
SEQ ID NO 3641  
LENGTH: 421  
TYPE: DNA  
ORGANISM: Homo sapien  
US-09-867-701-3641

Query Match 2.1%; Score 30; DB 10; Length 421;  
Best Local Similarity 100.0%; Pred. No. 8e-05;  
Matches 30; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1361 TCTTCCCTTAAAAAAAAAAAAAAAAAAAAA 1390  
DB 30 TCTTCCCTTAAAAAAAAAAAAAAAAAAAAA 1

RESULT 7  
US-09-770-445-884/c  
Sequence 884, Application US/09770445  
Patent No. US2002023281A1

GENERAL INFORMATION:  
APPLICANT: Gorlach, Jorn  
APPLICANT: An, Yong-Qiang  
APPLICANT: Hamilton, Carol M.  
APPLICANT: Price, Jennifer L.  
APPLICANT: Raines, Tracy M.  
APPLICANT: Yu, Yang  
APPLICANT: Rameaka, Joshua G.  
APPLICANT: Page, Amy  
APPLICANT: Matthew, Abraham V.  
APPLICANT: Ledford, Brooke L.  
APPLICANT: Moesner, Jeffrey P.  
APPLICANT: Haas, William David  
APPLICANT: Garcia, Carlos A.  
APPLICANT: Krickler, Maja  
APPLICANT: Slader, Ted  
APPLICANT: Davis, Keith R.  
APPLICANT: Allen, Keith  
APPLICANT: Hoffman, Neil  
APPLICANT: Hubman, Patrick

TITLE OF INVENTION: Expressed Sequences of Arabidopsis  
TITLE OF INVENTION: Chailana  
FILE REFERENCE: 2023US (PARA-012PRV)  
CURRENT APPLICATION NUMBER: US/09/770,445  
CURRENT FILING DATE: 2001-01-26  
PRIOR APPLICATION NUMBER: US 60/178,472  
PRIOR FILING DATE: 2000-01-27  
NUMBER OF SEQ ID NOS: 999  
SOFTWARE: FastSeq for Windows Version 4.0

SEQ ID NO 884.  
LENGTH: 777  
TYPE: DNA  
ORGANISM: Arabidopsis thaliana  
US-09-770-445-884

Query Match  
Best Local Similarity 100.0%; Pred. No. 7.9e-05;  
Matches 30; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1361 TCTTCCTTAAAAAAAAAAAAAAAAAAAA 1390  
DB 45 TCTTCCTTAAAAAAAAAAAAAAAAAAAA 16

RESULT 8  
US-10-174-590-569

Sequence 569, Application US/10174590  
Publication No. US20030008352A1

GENERAL INFORMATION:  
APPLICANT: Baker, Kevin P.  
APPLICANT: Chen, Jian  
APPLICANT: Desnoyers, Luc  
APPLICANT: Goddard, Audrey  
APPLICANT: Godowski, Paul J.  
APPLICANT: Gurney, Austin L.  
APPLICANT: Pan, James  
APPLICANT: Smith, Victoria  
APPLICANT: Watanabe, Colin K.  
APPLICANT: Wood, William I.  
APPLICANT: Zhang, Zemin  
TITLE OF INVENTION: SECRETED AND TRANSMEMBRANE POLYPEPTIDES AND NUCLEIC  
FILE REFERENCE: P3430R1C42  
CURRENT APPLICATION NUMBER: US/10/174,590  
CURRENT FILING DATE: 2002-06-18  
Prior application removed - See File Wrapper or Palm  
NUMBER OF SEQ ID NOS: 612  
SEQ ID NO 569  
LENGTH: 2457  
TYPE: DNA  
ORGANISM: Homo Sapien  
US-10-174-590-569

Query Match  
Best Local Similarity 100.0%; Pred. No. 7.7e-05;  
Matches 30; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1361 TCTTCCTTAAAAAAAAAAAAAAAAAAAA 1390  
DB 2423 TCTTCCTTAAAAAAAAAAAAAAAAAAAA 2452

RESULT 9  
US-10-176-758-569

Sequence 569, Application US/10176758  
Publication No. US20030008353A1

GENERAL INFORMATION:  
APPLICANT: Baker, Kevin P.  
APPLICANT: Chen, Jian  
APPLICANT: Desnoyers, Luc  
APPLICANT: Goddard, Audrey  
APPLICANT: Godowski, Paul J.  
APPLICANT: Gurney, Austin L.  
APPLICANT: Pan, James  
APPLICANT: Smith, Victoria  
APPLICANT: Watanabe, Colin K.  
APPLICANT: Wood, William I.  
APPLICANT: Zhang, Zemin  
TITLE OF INVENTION: SECRETED AND TRANSMEMBRANE POLYPEPTIDES AND NUCLEIC  
FILE REFERENCE: P3430R1C104  
CURRENT APPLICATION NUMBER: US/10/176,758

CURRENT FILING DATE: 2002-06-21  
Prior Application removed - See File Wrapper or Palm  
NUMBER OF SEQ ID NOS: 612  
SEQ ID NO 569  
LENGTH: 2457  
TYPE: DNA  
ORGANISM: Homo Sapien  
US-10-176-758-569

Query Match  
Best Local Similarity 100.0%; Pred. No. 7.7e-05;  
Matches 30; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1361 TCTTCCTTAAAAAAAAAAAAAAAAAAAA 1390  
DB 2423 TCTTCCTTAAAAAAAAAAAAAAAAAAAA 2452

RESULT 10  
US-10-175-737-569

Sequence 569, Application US/10175737  
Publication No. US20030013153A1

GENERAL INFORMATION:  
APPLICANT: Baker, Kevin P.  
APPLICANT: Chen, Jian  
APPLICANT: Desnoyers, Luc  
APPLICANT: Goddard, Audrey  
APPLICANT: Godowski, Paul J.  
APPLICANT: Gurney, Austin L.  
APPLICANT: Pan, James  
APPLICANT: Smith, Victoria  
APPLICANT: Watanabe, Colin K.  
APPLICANT: Wood, William I.  
APPLICANT: Zhang, Zemin  
TITLE OF INVENTION: SECRETED AND TRANSMEMBRANE POLYPEPTIDES AND NUCLEIC  
FILE REFERENCE: P3430R1C50  
CURRENT APPLICATION NUMBER: US/10/175,737  
CURRENT FILING DATE: 2002-06-19  
Prior Application removed - See File Wrapper or Palm  
NUMBER OF SEQ ID NOS: 612  
SEQ ID NO 569  
LENGTH: 2457  
TYPE: DNA  
ORGANISM: Homo Sapien  
US-10-175-737-569

Query Match  
Best Local Similarity 100.0%; Pred. No. 7.7e-05;  
Matches 30; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1361 TCTTCCTTAAAAAAAAAAAAAAAAAAAA 1390  
DB 2423 TCTTCCTTAAAAAAAAAAAAAAAAAAAA 2452

RESULT 11  
US-10-173-706-569

Sequence 569, Application US/10173706  
Publication No. US2003002293A1

GENERAL INFORMATION:  
APPLICANT: Baker, Kevin P.  
APPLICANT: Chen, Jian  
APPLICANT: Desnoyers, Luc  
APPLICANT: Goddard, Audrey  
APPLICANT: Godowski, Paul J.  
APPLICANT: Gurney, Austin L.  
APPLICANT: Pan, James  
APPLICANT: Smith, Victoria  
APPLICANT: Watanabe, Colin K.  
APPLICANT: Wood, William I.  
APPLICANT: Zhang, Zemin  
TITLE OF INVENTION: SECRETED AND TRANSMEMBRANE POLYPEPTIDES AND NUCLEIC



TITLE OF INVENTION: ACIDS ENCODING THE SAME  
FILE REFERENCE: P3430R1C7  
CURRENT APPLICATION NUMBER: US/10/173,706  
CURRENT FILING DATE: 2002-06-17  
Prior Application removed - See File Wrapper or Palm  
NUMBER OF SEQ ID NOS: 612  
SEQ ID NO 569  
LENGTH: 2457  
TYPE: DNA  
ORGANISM: Homo Sapien  
US-10-173-706-569

Query Match 2.1%; Score 30; DB 9; Length 2457;  
Best Local Similarity 100.0%; Pred. No. 7.7e-05;  
Matches 30; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1361 TCTTCCCTAATAAAAAAAAAAAAAAAAAA 1390  
DB 2423 TCTTCCCTAATAAAAAAAAAAAAAAAAAA 2452

## RESULT 12

US-10-175-738-569  
Sequence 569, Application US/10175738  
Publication No. US20030022294A1  
GENERAL INFORMATION:

APPLICANT: Baker, Kevin P.  
APPLICANT: Chen, Jian  
APPLICANT: Desnoyers, Luc  
APPLICANT: Goddard, Audrey  
APPLICANT: Godowski, Paul J.  
APPLICANT: Gurney, Austin L.  
APPLICANT: Pan, James  
APPLICANT: Smith, Victoria  
APPLICANT: Watanabe, Colin K.  
APPLICANT: Wood, William I.  
APPLICANT: Zhang, Zemin  
TITLE OF INVENTION: SECRETED AND TRANSMEMBRANE POLYPEPTIDES AND NUCLEIC  
FILE REFERENCE: P3430R1C45  
CURRENT APPLICATION NUMBER: US/10/175,738  
CURRENT FILING DATE: 2002-06-19  
Prior Application removed - See File Wrapper or Palm  
NUMBER OF SEQ ID NOS: 612  
SEQ ID NO 569  
LENGTH: 2457  
TYPE: DNA  
ORGANISM: Homo Sapien  
US-10-175-738-569

Query Match 2.1%; Score 30; DB 9; Length 2457;  
Best Local Similarity 100.0%; Pred. No. 7.7e-05;  
Matches 30; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1361 TCTTCCCTAATAAAAAAAAAAAAAAAAAA 1390  
DB 2423 TCTTCCCTAATAAAAAAAAAAAAAAAAAA 2452

## RESULT 13

US-10-175-752-569  
Sequence 569, Application US/10175752  
Publication No. US20030022295A1  
GENERAL INFORMATION:

APPLICANT: Baker, Kevin P.  
APPLICANT: Chen, Jian  
APPLICANT: Desnoyers, Luc  
APPLICANT: Goddard, Audrey  
APPLICANT: Godowski, Paul J.  
APPLICANT: Gurney, Austin L.  
APPLICANT: Pan, James  
APPLICANT: Smith, Victoria  
APPLICANT: Watanabe, Colin K.

APPLICANT: Wood, William I.  
APPLICANT: Zhang, Zemin  
TITLE OF INVENTION: SECRETED AND TRANSMEMBRANE POLYPEPTIDES AND NUCLEIC  
FILE REFERENCE: P3430R1C60  
CURRENT APPLICATION NUMBER: US/10/175,752  
CURRENT FILING DATE: 2002-06-19  
Prior Application removed - See File Wrapper or Palm  
NUMBER OF SEQ ID NOS: 612  
SEQ ID NO 569  
LENGTH: 2457  
TYPE: DNA  
ORGANISM: Homo Sapien  
US-10-175-752-569

Query Match 2.1%; Score 30; DB 9; Length 2457;  
Best Local Similarity 100.0%; Pred. No. 7.7e-05;  
Matches 30; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1361 TCTTCCCTAATAAAAAAAAAAAAAAAAAA 1390  
DB 2423 TCTTCCCTAATAAAAAAAAAAAAAAAAAA 2452

## RESULT 14

US-10-176-482-569  
Sequence 569, Application US/10176482  
Publication No. US20030022296A1  
GENERAL INFORMATION:

APPLICANT: Baker, Kevin P.  
APPLICANT: Chen, Jian  
APPLICANT: Desnoyers, Luc  
APPLICANT: Goddard, Audrey  
APPLICANT: Godowski, Paul J.  
APPLICANT: Gurney, Austin L.  
APPLICANT: Pan, James  
APPLICANT: Smith, Victoria  
APPLICANT: Watanabe, Colin K.  
APPLICANT: Wood, William I.  
APPLICANT: Zhang, Zemin  
TITLE OF INVENTION: SECRETED AND TRANSMEMBRANE POLYPEPTIDES AND NUCLEIC  
FILE REFERENCE: P3430R1C70  
CURRENT APPLICATION NUMBER: US/10/176,482  
CURRENT FILING DATE: 2002-06-20  
Prior Application removed - See File Wrapper or Palm  
NUMBER OF SEQ ID NOS: 612  
SEQ ID NO 569  
LENGTH: 2457  
TYPE: DNA  
ORGANISM: Homo Sapien  
US-10-176-482-569

Query Match 2.1%; Score 30; DB 9; Length 2457;  
Best Local Similarity 100.0%; Pred. No. 7.7e-05;  
Matches 30; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1361 TCTTCCCTAATAAAAAAAAAAAAAAAAAA 1390  
DB 2423 TCTTCCCTAATAAAAAAAAAAAAAAAAAA 2452

## RESULT 15

US-10-176-757-569  
Sequence 569, Application US/10176757  
Publication No. US20030022297A1  
GENERAL INFORMATION:

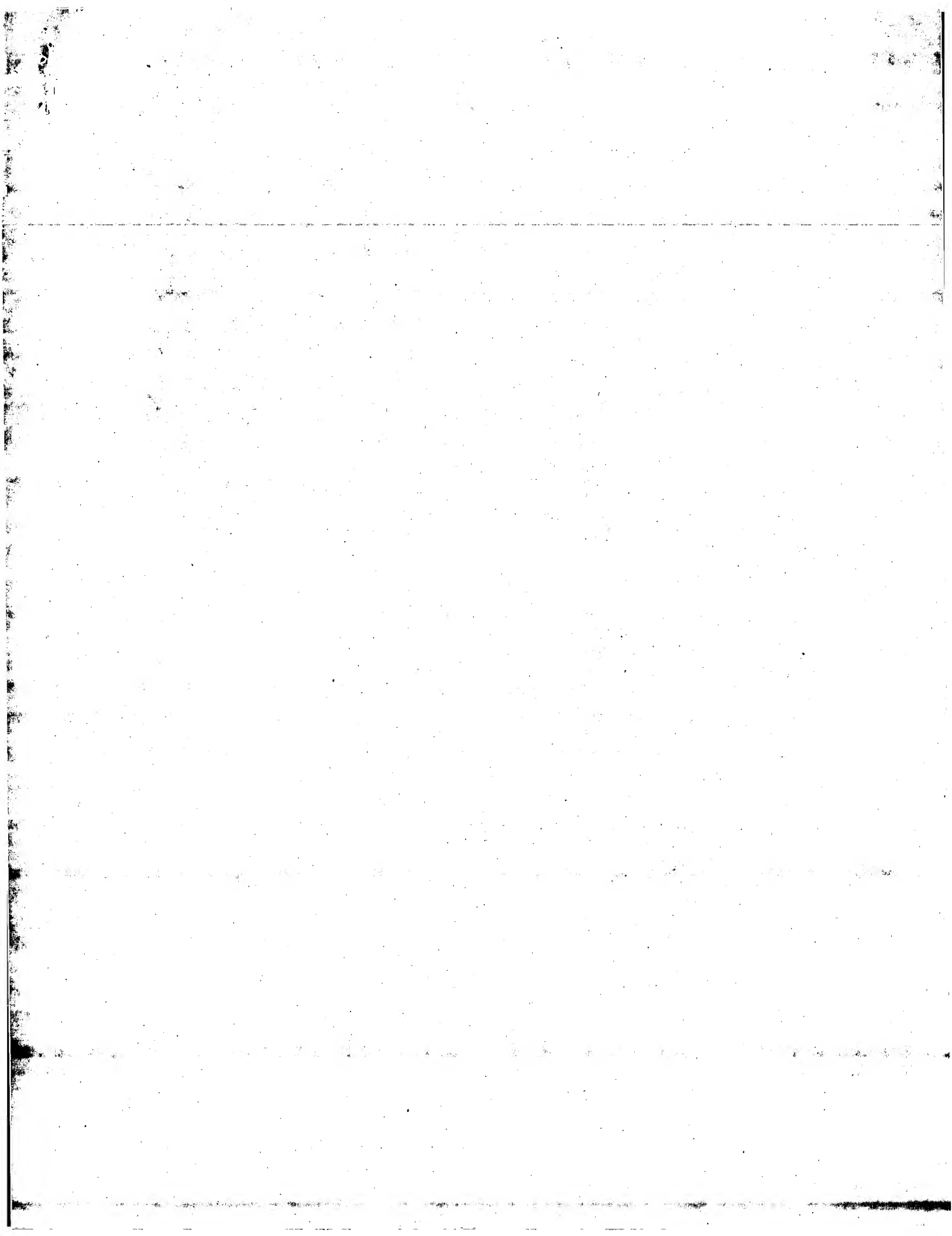
APPLICANT: Baker, Kevin P.  
APPLICANT: Chen, Jian  
APPLICANT: Desnoyers, Luc  
APPLICANT: Goddard, Audrey  
APPLICANT: Godowski, Paul J.  
APPLICANT: Gurney, Austin L.



APPLICANT: Pan,James  
 APPLICANT: Smith,Victoria  
 APPLICANT: Watanabe,Colin K.  
 APPLICANT: Wood,William I.  
 APPLICANT: Zhang,Zemin  
 TITLE OF INVENTION: SECRETED AND TRANSMEMBRANE POLYPEPTIDES AND NUCLEIC  
 FILE REFERENCE: P3430R1C86  
 CURRENT APPLICATION NUMBER: US/10/176,757  
 CURRENT FILING DATE: 2002-06-20  
 Prior Application removed - See File Wrapper or Palm  
 NUMBER OF SEQ ID NOS: 612  
 SEQ ID NO 569  
 LENGTH: 2457  
 TYPE: DNA  
 ORGANISM: Homo Sapien  
 US-10-176-757-569

Query Match 2.1%; Score 30; DB 9; Length 2457;  
 Best Local Similarity 100.0%; Pred. No. 7.7e-05;  
 Matches 30; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 QY 1361 TCTTCCTAAAAAAAAAAAAAAAAAAAA 1390  
 DB 2423 TCTTCCTAAAAAAAAAAAAAAAAAAAA 2452

Search completed: June 29, 2003, 06:40:51  
 Job time : 185 secs



















956	27	1.9	984	12	BS635590	BS635590 602582319
957	27	1.9	988	13	BM416548	BM416548 OP21634 M
958	27	1.9	1001	14	BQ729018	BQ729018 AGENCOURT
959	27	1.9	1006	14	BQ877143	BQ877143 AGENCOURT
960	27	1.9	1010	12	BF797153	BF797153 AGENCOURT
961	27	1.9	1013	14	BQ727354	BQ727354 AGENCOURT
962	27	1.9	1014	14	BQ950220	BQ950220 AGENCOURT
963	27	1.9	1016	17	CNS03BYV	AL255280 Tetradon
964	27	1.9	1025	12	BF034237	BF034237 601456031
965	27	1.9	1027	14	BQ844400	BQ844400 AGENCOURT
966	27	1.9	1040	14	BQ944642	BQ944642 AGENCOURT
967	27	1.9	1049	13	BM553565	BM553565 AGENCOURT
968	27	1.9	1051	12	BF168559	BF168559 601775462
969	27	1.9	1056	9	AL513931	AL513931 AL513931
970	27	1.9	1060	12	BF033752	BF033752 601454087
971	27	1.9	1063	13	BT734239	BT734239 603351348
972	27	1.9	1069	13	BG912561	BG912561 602806674
973	27	1.9	1076	12	BE962688	BE962688 601656062
974	27	1.9	1082	14	BQ930141	BQ930141 AGENCOURT
975	27	1.9	1085	13	BM564621	BM564621 AGENCOURT
976	27	1.9	1088	10	BE621067	BE621067 601493674
977	27	1.9	1090	14	BM907065	BM907065 AGENCOURT
978	27	1.9	1103	12	BG818955	BG818955 602779368
979	27	1.9	1105	13	BM455446	BM455446 AGENCOURT
980	27	1.9	1107	14	BQ946165	BQ946165 AGENCOURT
981	27	1.9	1110	14	BM928362	BM928362 AGENCOURT
982	27	1.9	1155	10	BE420976	BE420976 HMO04.E0
983	27	1.9	1162	13	BM476725	BM476725 AGENCOURT
984	27	1.9	1191	11	AY104119	AY104119 Zea mays
985	27	1.9	1211	14	BQ929879	BQ929879 AGENCOURT
986	27	1.9	1216	12	BG035441	BG035441 602325071
987	27	1.9	1217	13	BI688046	BI688046 603314622
988	27	1.9	1321	12	BF179216	BF179216 601807467
989	27	1.9	1328	14	BM467957	BM467957 AGENCOURT
990	27	1.9	1361	14	BM926544	BM926544 AGENCOURT
991	27	1.9	1430	14	BM805984	BM805984 AGENCOURT
992	27	1.9	1468	13	BM545669	BM545669 AGENCOURT
993	27	1.9	1583	12	BG252814	BG252814 602365383
994	27	1.9	1771	11	BE778572	BE778572 601466103
995	27	1.9	1771	11	AY103565	AY103565 Zea mays
996	27	1.9	1807	12	BE889423	BE889423 601512540
997	27	1.9	2020	11	AY103619	AY103619 Zea mays
998	27	1.9	2100	9	AT183456	AT183456 AU183456
999	26	1.9	50	13	BI865491	BI865491 ft23912.x
1000	26	1.9	59	9	AL802723	AL802723 AL802723

## ALIGNMENTS

RESULT 1	BM927766	1039 bp	mrna	linear	EST 12-MAR-2002
LOCUS	AGENCOURT_67292770	NIH_MGC_100	Homo sapiens	cdna	clone IMAGE:5797598
DEFINITION	5', mRNA sequence.				
ACCESSION	BM927766	GI:19378145			
VERSION	EST.				
KEYWORDS	human.				
SOURCE	Homo sapiens				
ORGANISM	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.				
REFERENCE	1 (bases 1 to 1039)				
AUTHORS	NIH-MGC <a href="http://mgc.nci.nih.gov/">http://mgc.nci.nih.gov/</a> .				
TITLE	National Institutes of Health, Mammalian Gene Collection (MGC)				
JOURNAL	Unpublished (1999)				
COMMENT	Contact: Robert Strausberg, Ph.D. Email: rgs@nih.gov Tissue Procurement: CGAP (Stanford) CDNA Library Preparation: Rubin Laboratory CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LNLN) DNA Sequencing by: Agencourt Bioscience Corporation Clone distribution: MGC clone distribution information can be				

## FEATURES

Found through the I.M.A.G.E. Consortium/LNLN at:  
http://image.lnl.gov  
Plate: L1CM2024 row: d column: 15  
High quality sequence stop: 655.  
Location/Qualifiers  
1..1039  
/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
/clone="IMAGE:5797598"  
/clone\_lib="NIH\_MGC\_100"  
/tissue\_type="hepatocellular carcinoma, cell line"  
/lab\_host="DH10B (phage-resistant)"  
/note="Organ: liver; Vector: pOTB7; Site: 1: XhoI; Site: 2: EcoRI; CDNA made by oligo-dt priming. Directionally cloned into EcoRI/XhoI sites using the following 5' adaptor: GGCAAGAG(G). Size-selected by 500bp for average insert size 1.8kb. Library constructed by Ling Hong in the laboratory of Gerald M. Rubin (University of California, Berkeley) using ZAP-CDNA synthesis kit (Stratagene) and Superscript II RT (Life Technologies). Note: this is a NIH\_MGC Library"

BASE COUNT 223 a 276 c 300 g 237 t 3 others  
ORIGIN

Query Match 42.0%; Score 587; DB 14; Length 1039;  
Best Local Similarity 99.8%; Pred. No. 0;  
Matches 637; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY	442	CAGGTGCGAGCTGCTTTTCAAGATGATGAGCCAGCTGCTCCATGCCAACA	501
DB	94	CAGGTGCGAGCTGCTTTTCAAGATGATGAGCCAGCTGCTCCATGCCAACA	153
QY	502	AACACCAATATGTAAGGCTGCTGCTATCATGAGGCTGCTGACAGCTGACT	561
DB	154	AACACCAATATGTAAGGCTGCTGCTATCATGAGGCTGCTGACAGCTGACT	213
QY	562	ATCCTCAGAGCTGAGCTGCTGCTGAGAGGAGTGAATGCACTGAGCACTGATGTC	621
DB	214	ATCCTCAGAGCTGAGCTGCTGCTGAGAGGAGTGAATGCACTGAGCACTGATGTC	273
QY	622	ACCTGGAACCCCTGAGCAAAAGTATACATCCAGACAGATGTATCCAGACAA	681
DB	274	ACCTGGAACCCCTGAGCAAAAGTATACATCCAGACAGATGTATCCAGACAA	333
QY	682	CGGCAATATGCAAAATGTAATGAGTTTACCACTAGATGAGACTGCTGAC	741
DB	334	CGGCAATATGCAAAATGTAATGAGTTTACCACTAGATGAGACTGCTGAC	393
QY	742	TCTATGTCAGAGATCATGAGGATATGATGCTCTCCAACTGAGCTGTAAGCAAG	801
DB	394	TCTATGTCAGAGATCATGAGGATATGATGCTCTCCAACTGAGCTGTAAGCAAG	453
QY	802	CTCAGGCTATGCTCCCACTGAGGAGCTGTGCCCCCTCTGCGAGAGGTTCCG	861
DB	454	CTCAGGCTATGCTCCCACTGAGGAGCTGTGCCCCCTCTGCGAGAGGTTCCG	513
QY	862	CCATCATCTGTTCAATAGTGAAGATAGTAAGGCTGCTGCTGCTGCTGCTGCA	921
DB	514	CCATCATCTGTTCAATAGTGAAGATAGTAAGGCTGCTGCTGCTGCTGCTGCA	573
QY	922	TGCGACAGAGCGGTGAGGAGCTGCTGAGGAGCAATCATGAGTGTCTCAGCT	981
DB	574	TGCGACAGAGCGGTGAGGAGCTGCTGAGGAGCAATCATGAGTGTCTCAGCT	633
QY	982	TAGGTGTGACAGAGACTTGGGGGAGTGTCTCCAGATGTGGTGTGTTACTGG	1041
DB	634	TAGGTGTGACAGAGACTTGGGGGAGTGTCTCCAGATGTGGTGTGTTACTGG	693
QY	1042	GGAGGCTATCTGACCTCCGACAGGAGCACTGCCA	1079
DB	694	GGAGGCTATCTGACCTCCGACAGGAGCACTGCCA	731

RESULT 2  
BG827019 776 bp mRNA linear EST 22-MAY-2001  
LOCUS 602749053F1 NIH\_MGC\_17 Homo sapiens cDNA clone IMAGE:4901869 5'  
DEFINITION mRNA sequence.  
ACCESSION BG827019  
VERSION BG827019.1 GI:14174606  
KEYWORDS EST.  
SOURCE human.  
ORGANISM Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.  
REFERENCE 1 (bases 1 to 776)  
NIH-MGC http://mgi.nci.nih.gov/  
NATIONAL INSTITUTES OF HEALTH, MAMMALIAN GENE COLLECTION (MGC)  
JOURNAL Unpublished (1999)  
COMMENT Contact: Robert Strausberg, Ph.D.  
Email: cgapbs-remail.nih.gov  
Tissue Procurement: ATCC  
CDNA Library Preparation: Ling Hong/Rubin Laboratory  
DNA Sequencing by: The I.M.A.G.E. Consortium (LNL)  
Clone distribution: MGC clone distribution information can be  
found through the I.M.A.G.E. Consortium/LNL at:  
http://image.lnl.gov  
Plate: LICM1797 row: 3 column: 14  
High quality sequence stop: 771.  
Location/Qualifiers  
1..776  
/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
/clone="IMAGE:4901869"  
/clone\_lib="NIH MGC 17"  
/tissue\_type="rhabdomyosarcoma"  
/lab\_host="DH10B (phage-resistant)"  
/note="Organ: muscle; Vector: pOTB7, Site: 1: EcoRI;  
Site: 2: XhoI; cDNA made by oligo-dT priming.  
Directionally cloned into EcoRI/XhoI sites using the  
following 5' adaptor: GCGACGAG(G). Site-selected >500bp  
for average insert size 1.8kb. Library constructed by  
Ling Hong in the laboratory of Gerald M. Rubin (University  
of California, Berkeley) using ZAP-CDNA synthesis kit  
(Stratagene) and Superscript II RT (Life Technologies)."  
BASE COUNT 158 a 228 c 222 g 168 t  
ORIGIN  
Query Match 40.5%; Score 566; DB 12; Length 776;  
Best Local Similarity 100.0%; Pred. No. 0;  
Matches 566; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
QY 442 CAGGTGGGACGCTGCTTTTCAAGACTGATGAGCCAGTGGTCCCTGATCCCAACA 501  
DB 202 CAGGTGGGACGCTGCTTTTCAAGACTGATGAGCCAGTGGTCCCTGATCCCAACA 261  
QY 502 AGACCAATATGTAAGGCTCTGGCTGACCTATCTGAGGGCTCGCTGACCACTGACT 561  
DB 262 AGACCAATATGTAAGGCTCTGGCTGACCTATCTGAGGGCTCGCTGACCACTGACT 321  
QY 562 ATCTCAGACGCTGGCTTCCCTGTGAGGAGTACTGCTGACCTGAGCAGCATGCTGC 621  
DB 322 ATCTCAGACGCTGGCTTCCCTGTGAGGAGTACTGCTGACCTGAGCAGCATGCTGC 381  
QY 622 ACCGTGGAAACCTCTGAGACAAAGCTAACATCCACAGACAGAGATGTGACGAGCAAA 681  
DB 382 ACCGTGGAAACCTCTGAGACAAAGCTAACATCCACAGACAGAGATGTGACGAGCAAA 441  
QY 682 CGTGCAATATGCAATATGTAATATGTAATATGTAATATGTAATATGTAATATGTA 741  
DB 442 CGTGCAATATGCAATATGTAATATGTAATATGTAATATGTAATATGTAATATGTA 501  
QY 742 TCTTAGTCCAGGATATGAGGGGTATGATGCTCTTCCAACTGTGGGCTGTAGCAAG 801  
DB 502 TCTTAGTCCAGGATATGAGGGGTATGATGCTCTTCCAACTGTGGGCTGTAGCAAG 561

QY 802 CTCAGGCTAGTCTCCCACTGAGGGGCTGGCCCTCTGAGACGCTTCCGTGGCAGCC 861  
DB 562 CTCAGGCTAGTCTCCCACTGAGGGGCTGGCCCTCTGAGACGCTTCCGTGGCAGCC 621  
QY 862 CCATCACTGTGTTCAATAGTGTAGAGATGAGTAAAGCCCTCTGCTGCTGCTGACA 921  
DB 622 CCATCACTGTGTTCAATAGTGTAGAGATGAGTAAAGCCCTCTGCTGCTGCTGACA 681  
QY 922 TGCACAGAGGCGGGGGGCTGCTGGGACATCCATCCGAGTGTCTCTAGCT 981  
DB 682 TGCACAGAGGCGGGGGGCTGCTGGGACATCCGAGTGTCTCTAGCT 741  
QY 982 TAGGTCTGACAGAGACTTGGCGGG 1007  
DB 742 TAGGTCTGACAGAGACTTGGCGGG 767  
RESULT 3  
BQ013425/c 520 bp mRNA linear EST 26-MAR-2002  
LOCUS UI-1-BC1p-ayx-c-01-0-UI s1 NCI CGAP\_P13 Homo sapiens cDNA clone  
DEFINITION UI-1-BC1p-ayx-c-01-0-UI 3', mRNA sequence.  
ACCESSION BQ013425  
VERSION BQ013425.1 GI:19738326  
KEYWORDS EST.  
SOURCE human.  
ORGANISM Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.  
REFERENCE 1 (bases 1 to 520)  
NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.  
NATIONAL CANCER INSTITUTE, CANCER GENOME ANATOMY PROJECT (CGAP),  
Tumor Gene Index  
Unpublished (1997)  
COMMENT Contact: Robert Strausberg, Ph.D.  
Email: cgapbs-remail.nih.gov  
Tissue Procurement: Dr. Steven Brown  
CDNA Library Preparation: Dr. M. Bento Soares, University of Iowa  
DNA Sequencing by: Dr. M. Bento Soares, University of Iowa  
Clone Distribution: Clone distribution information can be obtained  
from Dr. M. Bento Soares, bento-soares@uiowa.edu  
The following repetitive elements were found in this cDNA  
sequence: 44-82, >AT rich#low\_complexity  
Seq primer: M13 FORWARD  
POLYA=yes.  
Location/Qualifiers  
1..520  
/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
/clone="UI-1-BC1p-ayx-c-01-0-UI"  
/clone\_lib="NCI CGAP\_P13"  
/tissue\_type="Placenta"  
/dev stage="8-9 weeks"  
/lab\_host="DH10B (Life Technologies)"  
/note="Organ: Placenta; Vector: pRTT3-Pac (Pharmacia) with  
a modified polylinker; Site: 1: EcoRI, Site: 2: Not I;  
NCI CGAP\_P13 is a subcloned cDNA library constructed  
according to Bonaldo, Lennon and Soares, Genome Research,  
6:791-806, 1996. First strand cDNA synthesis was primed  
with an oligo-dT primer containing a Not I site. Double  
stranded cDNA was ligated to an EcoRI adaptor, digested  
with Not I, and cloned directionally into pRTT3-Pac  
vector. The oligonucleotide used to prime the synthesis of  
first-strand cDNA contains a library tag sequence that is  
located between the Not I site and the (dT)18 tail. The  
sequence tags for this library are GA, AGGA. For  
additional information, contact: Bento Soares,  
bento-soares@uiowa.edu  
TAG\_LIB=UI-1-BC1p  
TAG\_TISSUE=Placenta human 8 week  
TAG\_SEO=GA"

BASE COUNT 127 a 147 c 115 g 131 t  
 ORIGIN  
 Query Match 37.2%; Score 520; DB 14; Length 520;  
 Best Local Similarity 100.0%; Pred. No. 0;  
 Matches 520; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 867 ACTGTGTTCAATAGTGTGAGATGTAGCTAAACCCCTGCTGCTGCTGACATGCCA 926  
 DB 520 ACTGTGTTCAATAGTGTGAGATGTAGCTAAACCCCTGCTGCTGCTGACATGCCA 461

QY 927 CACAGAGCGGTGGGGGCTGGTGGGACATTCATGTGGAGTGTCTCTCAGCTTAGGT 986  
 DB 460 CACAGAGCGGTGGGGGCTGGTGGGACATTCATGTGGAGTGTCTCTCAGCTTAGGT 401

QY 987 CTGAGACGAGACTGTGGCGGGGATGCTCCAGAGATGTGGTGAATTCGTACTCTGGGAGG 1046  
 DB 400 CTGAGACGAGACTGTGGCGGGGATGCTCCAGAGATGTGGTGAATTCGTACTCTGGGAGG 341

QY 1047 CTATCTGTACTCTCCGACAGGGGACACTCCAGGCCAGGCCAGGGGTCAAGGGGAGAG 1106  
 DB 340 CTATCTGTACTCTCCGACAGGGGACACTCCAGGCCAGGCCAGGGGTCAAGGGGAGAG 281

QY 1107 TGACACCTCAGATGAGCCCAAGACTGGGGGTCAAGGAGCAGGTGTGGTTGAGCCAGGAC 1166  
 DB 280 TGACACCTCAGATGAGCCCAAGACTGGGGGTCAAGGAGCAGGTGTGGTTGAGCCAGGAC 221

QY 1167 CTGGGGCGGGGGGTGGGGGCGGGGCTTCTGCTCATTTTCTTCAATGAAAGCTCAAA 1226  
 DB 220 CTGGGGCGGGGGGTGGGGGCGGGGCTTCTGCTCATTTTCTTCAATGAAAGCTCAAA 161

QY 1227 GCAGCCAAACCGAGGCTTTTCCCTTCTGAGTTGAATATCCAGAACTTTTGTACTT 1286  
 DB 160 GCAGCCAAACCGAGGCTTTTCCCTTCTGAGTTGAATATCCAGAACTTTTGTACTT 101

QY 1287 CTGTGTGTTAAATTTTATTTTATTTTGTAAATAATTAATTAATTAATTAATTAAT 1346  
 DB 100 CTGTGTGTTAAATTTTATTTTATTTTGTAAATAATTAATTAATTAATTAATTAAT 41

QY 1347 GTTTCACAGCAAACTCTCCCTAAATAAAAAAAAAAAAAA 1386  
 DB 40 GTTTCACAGCAAACTCTCCCTAAATAAAAAAAAAAAAAA 1

RESULT 4  
 A1669816/c 529 bp mRNA linear EST 15-DEC-1999  
 LOCUS tuj1f08.x1 NCI CGAP\_P128 Homo sapiens cDNA clone IMAGE:2252679 3',  
 DEFINITION mRNA sequence.  
 ACCESSION A1669816  
 VERSION A1669816.1 GI:4834590  
 KEYWORDS EST.  
 SOURCE human.  
 ORGANISM Homo sapiens  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.  
 REFERENCE 1 (bases 1 to 529)  
 AUTHORS NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.  
 TITLE National Cancer Institute, Cancer Genome Anatomy Project (CGAP),  
 Tumor Gene Index  
 JOURNAL Unpublished (1997)  
 COMMENT Contact: Robert Strausberg, Ph.D.  
 Email: cgapbs-remail.nih.gov  
 Tissue Procurement: Michael J. Brownstein, M.D., Ph.D., Michael R.  
 Emmert-Buck, M.D., Ph.D.  
 cDNA Library Preparation: M. Bento Soares, Ph.D.  
 cDNA Library Arrayed by: Greg Lennon, Ph.D.  
 DNA Sequencing by: Washington University Genome Sequencing Center  
 Clone distribution: NCI-CGAP clone distribution information can be  
 found through the I.M.A.G.E. Consortium/UMLS at:  
 www-bio.lnli.gov/bbrp/image/image.html  
 Insert length: 912 Std Error: 0.00  
 Seq primer: -40up from Gibco

FEATURES  
 source  
 High quality sequence stop: 455.  
 Location/Qualifiers  
 1..529  
 /organism="Homo sapiens"  
 /db\_xref="taxon:9606"  
 /clone="IMAGE:2252679"  
 /clone\_id="NCI\_CGAP\_P128"  
 /sex="male"  
 /dev\_stage="adult"  
 /lab\_host="DH10B"  
 /note="Organ: prostate; Vector: pTR73D-Pac (Pharmacia)  
 with a modified polylinker; Plasmid DNA from the  
 normalized library NCI CGAP\_P128 was prepared, and ss  
 circles were made in vitro. Following HAP purification,  
 this DNA was used as tracer in a subtractive hybridization  
 reaction. The driver was PCR-amplified cDNAs from a pool  
 of 5,000 clones made from the same library (clonoids  
 985608-986759, 1101192-1101959, and 1217928-1220615).  
 Subtraction by Bento Soares and M. Fatima Bonaldo."

BASE COUNT 131 a 159 c 125 g 113 t  
 ORIGIN  
 Query Match 37.0%; Score 517; DB 9; Length 529;  
 Best Local Similarity 100.0%; Pred. No. 0;  
 Matches 517; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 840 TGGAGCGTTCGGTGGGAGAGCCCATCACTGTGTTCAATAGTGAATGAGCTAAAG 899  
 DB 517 TGGAGCGTTCGGTGGGAGAGCCCATCACTGTGTTCAATAGTGAATGAGCTAAAG 458

QY 900 CCCCTGCTGCTCTCTGACATGACAGCAGAGCGGTGGGGCTGCGTGGGAGCAATCC 959  
 DB 457 CCCCTGCTGCTCTCTGACATGACAGCAGAGCGGTGGGGCTGCGTGGGAGCAATCC 398

QY 960 ATCGTAGAGTTCCTCTCAGCTTAGGTGAGCAGAGACTTGGCGGGGAGATGCTCCAG 1019  
 DB 397 ATCGTAGAGTTCCTCTCAGCTTAGGTGAGCAGAGACTTGGCGGGGAGATGCTCCAG 338

QY 1020 ATGTGGTGATTCCTGACCTGGGAGGCTATCTGACCTCCCGACAGGGGACATCCCA 1079  
 DB 337 ATGTGGTGATTCCTGACCTGGGAGGCTATCTGACCTCCCGACAGGGGACATCCCA 278

QY 1080 GGCAGCCAGGAGGCTCAGGGGAGAGGTGACACACTCAGCATGAGCCAAAGCTGGGGTCA 1139  
 DB 277 GGCAGCCAGGAGGCTCAGGGGAGAGGTGACACACTCAGCATGAGCCAAAGCTGGGGTCA 218

QY 1140 GGGAGCAGGTGTGTTTGAGCCAGACCTGGGGGCGGGGCTTCTCTGCC 1199  
 DB 217 GGGAGCAGGTGTGTTTGAGCCAGACCTGGGGGCGGGGCTTCTCTGCC 158

QY 1200 TCATTTGCTTTCAATGAAGCTCAAAAGCAGCCAAACAGGCTTTCCCTTCTCGAG 1259  
 DB 157 TCATTTGCTTTCAATGAAGCTCAAAAGCAGCCAAACAGGCTTTCCCTTCTCGAG 98

QY 1260 TTGAATATCCAGAACTTTTGTACTCTTGTGTTGTTAAATTTTGTAAATAA 1319  
 DB 97 TTGAATATCCAGAACTTTTGTACTCTTGTGTTGTTAAATTTTGTAAATAA 38

QY 1320 TAAATATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 1356  
 DB 37 TAAATATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 1

RESULT 5  
 BE250262 770 bp mRNA linear EST 13-JUL-2000  
 LOCUS BE250262  
 DEFINITION 600943273F1 NIH\_MGC\_17 Homo sapiens cDNA clone IMAGE:2959770 5',  
 mRNA sequence.  
 ACCESSION BE250262  
 VERSION BE250262.1 GI:9120370  
 KEYWORDS EST.  
 SOURCE human.  
 ORGANISM Homo sapiens

REFERENCE 1 (bases 1 to 770)  
 AUTHORS NIH-MGC http://mgc.nci.nih.gov/  
 TITLE National Institutes of Health, Mammalian Gene Collection (MGC)  
 JOURNAL Unpublished (1999)  
 COMMENT Contact: Robert Strausberg, Ph.D.  
 Email: cgaabs-remail.nih.gov  
 Tissue Procurement: ATCC  
 CDNA Library Preparation: Ling Hong/Rubin Laboratory  
 CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LNL)  
 DNA Sequencing by: Incyte Genomics, Inc.  
 Clone distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/LNL at: image.lnl.gov  
 Plate: LHCMS2 row: a column: 19  
 High quality sequence stop: 730.

## FEATURES

Location/Qualifiers  
 1..770  
 /organism="Homo sapiens"  
 /db\_xref="taxon:9606"  
 /clone="IMAGE:2959770"  
 /clone.lib="NIH MGC 17"  
 /tissue\_type="rhodomyosarcoma"  
 /lab\_host="DH10B (phage-resistant)"  
 /note="Organ: muscle; Vector: pOTB7; Site: 1: EcoRI; Site 2: XhoI; CDNA made by oligo-dT priming. Directionally cloned into EcoRI/XhoI sites using the following 5' adaptor: GGCACGAG(C) Size-selected >500bp for average insert size 1.8kb. Library constructed by Ling Hong in the laboratory of Gerald M. Rubin (University of California, Berkeley) using ZAP-CDNA synthesis kit (Stratagene) and Superscript II RT (Life Technologies)."  
 157 a 219 c 226 g 168 t

BASE COUNT 157 a 219 c 226 g 168 t  
 ORIGIN  
 Query Match 36.8%; Score 514; DB 10; Length 770;  
 Best Local Similarity 100.0%; Pred. No. 0;  
 Matches 514; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

442 CAGGTGGGAGGCTGCTTTTCAAGACGTGATGAGGACCAAGTGGTCCCTGATCCCAACA 501  
 Db 185 CAGGTGGGAGGCTGCTTTTCAAGACGTGATGAGGACCAAGTGGTCCCTGATCCCAACA 244  
 QY 502 AGACCAATATGTAAGGCTCTGCTGATCTATCTGAGGGCTCGGCTGACCAAGTACT 561  
 Db 245 AGACCAATATGTAAGGCTCTGCTGATCTATCTGAGGGCTCGGCTGACCAAGTACT 304  
 QY 562 ATCTCAGACAGCTGGGCTTGCCTGTGAGGAGAGTGAATTGCACTGGCAGACATGATGTC 621  
 Db 305 ATCTCAGACAGCTGGGCTTGCCTGTGAGGAGAGTGAATTGCACTGGCAGACATGATGTC 364  
 QY 622 ACCTGGGAACCCCTGAGACAAAGCTAACATCCCAACAGACAGATGTGACCCGACAA 681  
 Db 365 ACCTGGGAACCCCTGAGACAAAGCTAACATCCCAACAGACAGATGTGACCCGACAA 424  
 QY 682 CGTGCAATATGTAAGGCTTAAATGTAGTTTACAGCCTGATGAGGACTGCTGGC 741  
 Db 425 CGTGCAATATGTAAGGCTTAAATGTAGTTTACAGCCTGATGAGGACTGCTGGC 484  
 QY 742 TCTTACTCAGAGATCATGGGGGTATGACTGCTTCCAACTGTGGGGCTGTAAACAG 801  
 Db 485 TCTTACTCAGAGATCATGGGGGTATGACTGCTTCCAACTGTGGGGCTGTAAACAG 544  
 QY 802 CTCAGAGTATGCTCCCACTGGGGGCTGTGGCCCTCCCTGGAGACGTTCCGTGGGACGC 861  
 Db 545 CTCAGAGTATGCTCCCACTGGGGGCTGTGGCCCTCCCTGGAGACGTTCCGTGGGACGC 604  
 QY 862 CCATCACTGTGTTCAATAGTGTGAGATGTAAAGCCCTGCTGCTGCTGCTGACA 921  
 Db 605 CCATCACTGTGTTCAATAGTGTGAGATGTAAAGCCCTGCTGCTGCTGCTGACA 664  
 QY 922 TGCACAGACAGCGGTGGGGGCTGCGTGGGACA 955

Db 665 TGCACAGACAGCGGTGGGGGCTGCGTGGGACA 698

RESULT 6  
 LOCUS BM671384/c 773 bp mRNA linear EST 27-FEB-2002  
 DEFINITION UI-E-CK1-afk-m-06-0-UI.s2 UI-E-CK1 Homo sapiens CDNA clone  
 ACCESSION UI-E-CK1-afk-m-06-0-UI.3', mRNA sequence.  
 VERSION BM671384  
 KEYWORDS BM671384.1 GI:18981282  
 EST.

SOURCE human.  
 ORGANISM Homo sapiens  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.  
 TITLE 1 (bases 1 to 773)  
 AUTHORS Ronald, M.F., Lennon, G. and Soares, M.B.  
 Normalization and subtraction: two approaches to facilitate gene discovery

JOURNAL Genome Res. 6 (9), 791-806 (1996)  
 MEDLINE 97044477  
 COMMENT Contact: Soares, MB  
 Program for Rat Gene Discovery and Mapping  
 University of Iowa  
 451 Eckstein Medical Research Building Iowa City, IA 52242, USA  
 Tel: 319 335 8250  
 Fax: 319 335 9565  
 Email: msoares@blue.weeg.uiowa.edu

Tissue Procurement: Dr. Gregg Hageman  
 CDNA Library Preparation: Dr. M. Bento Soares, University of Iowa  
 CDNA Library Arrayed by: Dr. M. Bento Soares, University of Iowa  
 DNA Sequencing by: Dr. M. Bento Soares, University of Iowa  
 Clone Distribution: Researchers may obtain clones from Research Genetics (www.resgen.com)  
 The following repetitive elements were found in this CDNA sequence: 1-59, >(TAAA)n#Simple\_repeat (matched complement)  
 Seq primer: M13 Forward  
 POLYA=yes.

## FEATURES

Location/Qualifiers  
 1..773  
 /organism="Homo sapiens"  
 /db\_xref="taxon:9606"  
 /clone="UI-E-CK1-afk-m-06-0-UI"  
 /clone.lib="UI-E-CK1"  
 /tissue\_type="Retina Foveal and Macular"  
 /dev\_stage="adult"  
 /lab\_host="DH10B (Life Technologies) (T1 phage resistant)"  
 /note="Organ: eye; Vector: pT73-Pac (Pharmacia) with a modified polylinker; Site 1: EcoR I; Site 2: Not I; UI-E-CK1 is a normalized CDNA library containing the following tissue(s): Retina Foveal and Macular. The library was constructed according to Bonaldo, Lennon and Soares, Genome Research, 6:791-806, 1996. First strand CDNA synthesis was primed with an oligo-dT primer containing a Not I site. Double stranded CDNA was ligated to an EcoR I adaptor, digested with Not I, and cloned directionally into pT73-Pac vector. The oligonucleotide used to prime the synthesis that is located between the Not I site and the (dT)18 tail. The sequence tag for this library is GTCC. This library was created for the program, Gene Discovery in the Visual System, supported by National Eye Institute (NEI).  
 TAG\_LIB=UI-E-CK1  
 TAG\_TISSUE=Foveal and Macular Retina  
 TAG\_SEQ=GTCC"

BASE COUNT 181 a 221 c 180 g 189 t 2 others  
 ORIGIN  
 Query Match 36.7%; Score 513; DB 14; Length 773;  
 Best Local Similarity 99.8%; Pred. No. 0;  
 Matches 563; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 780 AACCTGTGGGCTGTAGCAAGCTCAGGCTAGTCTCCCACTGGGGGCTGTGCCCCCTCC 839  
 DB 595 AACCTGTGGGCTGTAGCAAGCTCAGGCTAGTCTCCCACTGGGGGCTGTGCCCCCTCC 536  
 QY 840 TGGGACGGTCCGTGGGACAGCCCATCACTGTGTCAATAGTGTAGTAAG 899  
 DB 535 TGGGACGGTCCGTGGGACAGCCCATCACTGTGTCAATAGTGTAGTAAG 476  
 QY 900 CCCCTGTGCTGTGTGTCATGCAATGCAAGCAGCGGTGGGGGCTGCTGGGGCAATCC 959  
 DB 475 CCCCTGTGCTGTGTGTCATGCAATGCAAGCAGCGGTGGGGGCTGCTGGGGCAATCC 416  
 QY 960 ATGTGAGAGTCTCTCAGCTTGTGAGTGTGAGAGAGCTTGGGGGGGATGCTCCAG 1019  
 DB 415 ATGTGAGAGTCTCTCAGCTTGTGAGTGTGAGAGAGCTTGGGGGGGATGCTCCAG 356  
 QY 1020 ATGTGAGTGTCTGTACCTGAGGAGGCTATCTCTGACCTCCGACAGGGGACATCCCA 1079  
 DB 355 ATGTGAGTGTCTGTACCTGAGGAGGCTATCTCTGACCTCCGACAGGGGACATCCCA 296  
 QY 1080 GGGCAGCCAGGGGCTCAGGGGACAGAGGTGACACCTCAGATGAGCCAGACTGGGGTCA 1139  
 DB 295 GGGCAGCCAGGGGCTCAGGGGACAGAGGTGACACCTCAGATGAGCCAGACTGGGGTCA 236  
 QY 1140 GGGACAGAGTGTGTGAGCAGGACCTGGGGGCGGGGGGCGGGGCTTCTTGGCC 1199  
 DB 235 GGGACAGAGTGTGTGAGCAGGACCTGGGGGCGGGGGGCGGGGCTTCTTGGCC 176  
 QY 1200 TCATTGCTTTCATGAAAGCCCTCAAGCAGCCAAACAGAGGCTTCCCTTCTCGAG 1259  
 DB 175 TCATTGCTTTCATGAAAGCCCTCAAGCAGCCAAACAGAGGCTTCCCTTCTCGAG 116  
 QY 1260 TTGGAATATCCAGATCTTTTGTACTTCTTGTGTGTTAAATTTTGTAAATAA 1319  
 DB 115 TTGGAATATCCAGATCTTTTGTACTTCTTGTGTGTTAAATTTTGTAAATAA 56  
 QY 1320 TAAATTAATTAATTAATTAATTAAT 1343  
 DB 55 TAAATTAATTAATTAATTAATTAAT 32

RESULT 7  
 BM668538/c  
 LOCUS 610 bp mRNA linear EST 27-FEB-2002  
 DEFINITION UI-E-CK1-afm-c-23-0-UI.s2 UI-E-CK1 Homo sapiens cDNA clone  
 UI-E-CK1-afm-c-23-0-UI 3', mRNA sequence.  
 ACCESSION BM668538  
 VERSION BM668538.1 GI:18976369  
 KEYWORDS EST.  
 SOURCE human.  
 ORGANISM Homo sapiens  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.  
 REFERENCE 1 (bases 1 to 610)  
 AUTHORS Bonaldo, M.F., Lennon, G. and Soares, M.B.  
 TITLE Normalization and subtraction: two approaches to facilitate gene  
 discovery  
 JOURNAL Genome Res. 6 (9), 791-806 (1996)  
 MEDLINE 97044477  
 COMMENT Contact: Soares, MB  
 Program for Rat Gene Discovery and Mapping  
 University of Iowa  
 451 Eckstein Medical Research Building Iowa City, IA 52242, USA  
 Tel: 319 335 8250  
 Fax: 319 335 9565  
 Email: msoares@iue.wiueg.iowa.edu  
 Tissue Procurement: Dr. Gregg Hageman  
 cDNA Library preparation: Dr. M. Bento Soares, University of Iowa  
 cDNA Library Arrayed by: Dr. M. Bento Soares, University of Iowa  
 DNA Sequencing by: Dr. M. Bento Soares, University of Iowa  
 Clone Distribution: Researchers may obtain clones from Research  
 Genetics (www.resgen.com)  
 The following repetitive elements were found in this cDNA

sequence: 32-80, >AT rich#low\_complexity  
 Seq primer: M13 Forward  
 Polya=Yes  
 FEATURES  
 source  
 Location/Qualifiers  
 1..610  
 /organism="Homo sapiens"  
 /db\_xref="taxon:9606"  
 /clone="UI-E-CK1-afm-c-23-0-UI"  
 /clone\_lib="UI-E-CK1"  
 /tissue\_type="Retina Foveal and Macular"  
 /dev\_stage="adult"  
 /lab\_host="DH10B (Life Technologies) (T1 phage resistant)"  
 /note="Organ: eye; Vector: pT73-Pac (Pharmacia) with a  
 modified polylinker; Site 1: EcoR I; Site 2: Not I;  
 UI-E-CK1 is a normalized cDNA library containing the  
 following tissue(s): Retina Foveal and Macular. The  
 library was constructed according to Bonaldo, Lennon and  
 Soares, Genome Research, 6:791-806, 1996. First strand  
 cDNA synthesis was primed with an oligo-dT primer  
 containing a Not I site. Double stranded cDNA was ligated  
 to an EcoR I adaptor, digested with Not I, and cloned  
 directionally into pT73-Pac vector. The oligonucleotide  
 used to prime the synthesis of first-strand cDNA contains  
 a library tag sequence that is located between the Not I  
 site and the (dr)18 tail. The sequence tag for this  
 library is GTCC. This library was created for the program,  
 Gene Discovery in the Visual System, supported by National  
 Eye Institute (NEI).  
 TAG\_LIB=UI-E-CK1  
 TAG\_TISSUE=Foveal and Macular Retina  
 TAG\_SEQ=GTCC"  
 BASE COUNT 145 a 174 c 147 g 143 t 1 others  
 ORIGIN  
 Query Match 35.3%; Score 493; DB 13; Length 610;  
 Best Local Similarity 99.7%; Pred. No. 0;  
 Matches 593; Conservative 0; Mismatches 2; Indels 0; Gaps 0;  
 QY 765 TATGACTGCTTCCAAACCTGTGGGCTGTAGCAAGCTCAGGCTAGTCTCCCACTGG 824  
 DB 610 TATGACTGCTTCCAAACCTGTGGGCTGTAGCAAGCTCAGGCTAGTCTCCCACTGG 551  
 QY 825 GCGTGTGCCCCCTCCCTGGGACCGGTTCCGTGGGACGCCCATCATCTGTGTCAATAGTGTG 884  
 DB 550 GCGTGTGCCCCCTCCCTGGGACCGGTTCCGTGGGACGCCCATCATCTGTGTCAATAGTGTG 491  
 QY 885 AGAATGTACTTAAAGCCCTCTGCTGCTGTGCTGCACATGCCCACACAGGCGGTGGGGGT 944  
 DB 490 AGAATGTACTTAAAGCCCTCTGCTGCTGTGCTGCACATGCCCACACAGGCGGTGGGGGT 431  
 QY 945 GCGTGGGACATTCATGCTGAGAGTGTCTCTCAGCTTGTGTGAGAGAGACTTGGC 1004  
 DB 430 GCGTGGGACATTCATGCTGAGAGTGTCTCTCAGCTTGTGTGAGAGAGACTTGGC 371  
 QY 1005 GGGGAGTGTCCAGAGATGTGGTGTGTACTCTTGGGGAGGCTATCTTGAACCTCCCA 1064  
 DB 370 GGGGAGTGTCCAGAGATGTGGTGTGTACTCTTGGGGAGGCTATCTTGAACCTCCCA 311  
 QY 1065 CAGGGACACTCCAGGCGCAGCCCGAGGGGTCAAGGGGCAAGGTGACACACTCAGACATGAG 1124  
 DB 310 CAGGGACACTCCAGGCGCAGCCCGAGGGGTCAAGGGGCAAGGTGACACACTCAGACATGAG 251  
 QY 1125 CCAGAGCTGGGGTCAAGGAGCAGAGTGTGTTGAGCCAGAGCTTGGGGGCGGGGTGGGGC 1184  
 DB 250 CCAGAGCTGGGGTCAAGGAGCAGAGTGTGTTGAGCCAGAGCTTGGGGGCGGGGTGGGGC 191  
 QY 1185 CCGGGGCTTTTGTGCTCATTTGCTTCAATGAAGCTTCAAGCAGCCAAACACAGGCTT 1244  
 DB 190 CCGGGGCTTTTGTGCTCATTTGCTTCAATGAAGCTTCAAGCAGCCAAACACAGGCTT 131  
 QY 1245 TCCCTCTCTCGAGTGTGAATTCAGAAATCTTTTGTACTCTTGTGGTTAAATTTGT 1304  
 DB 130 TCCCTCTCTCGAGTGTGAATTCAGAAATCTTTTGTACTCTTGTGGTTAAATTTGT 71



QY 1305 TATTTTGTAAATAAATAAATAGTAAATGATGTTTTCACAGCAAA 1359  
 DB 70 TATTTTGTAAATAAATAAATAAATAGTAAATGATGTTTTCACAGCAAA 16

RESULT 8  
 BM142660/c 543 bp mRNA linear EST 29-NOV-2001  
 LOCUS 192602.x1 Human Fetal Pancreas 1B Homo sapiens cDNA 3', mRNA  
 DEFINITION sequence.  
 ACCESSION BM142660  
 VERSION BM142660.1 GI:17152727  
 KEYWORDS EST.  
 SOURCE human.  
 ORGANISM Homo sapiens  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.  
 REFERENCE 1 (bases 1 to 543)  
 AUTHORS Melton, D., Brown, J., Kenty, G., Permut, A., Lee, C., Kaestner, R.,  
 Lemishka, I., Searce, M., Brestelli, J., Gradwohl, G., Clifton, S.,  
 Hillier, L., Marra, M., Pape, D., Wylie, T., Martin, J., Biststein, A.,  
 Schmitt, A., Theising, B., Ritzer, E., Komko, I., Bennett, J., Cardenas,  
 M., Gibbons, W., McCann, R., Cole, R., Tsagaris, R., Williams, T.,  
 Jackson, Y., and Bowers, Y.  
 Endocrine Pancreas Consortium  
 Unpublished (2000)  
 CONTACT: Douglas Melton, Klaus H. Kaestner, & Hiroshi Inoue  
 Endocrine Pancreas Consortium  
 Harvard University, Howard Hughes Medical Institute  
 Dept of Molecular and Cellular Biology, 7 Divinity Ave, Cambridge,  
 MA 02138  
 Tel: 617-495-1812  
 Fax: 617-495-8557  
 Email: dmelton@chp.harvard.edu  
 Library was constructed by Dr. Hiroshi Inoue DNA sequencing by:  
 Washington University Genome Sequencing Center For information on  
 obtaining a clone please contact: Dr. Hiroshi Inoue  
 (hinoue@imgate.wustl.edu)  
 Seq primer: -40UP from Gibco  
 High quality sequence stop: 445.  
 Location/Qualifiers  
 1. 543  
 /organism="Homo sapiens"  
 /db\_xref="taxon:9606"  
 /clone\_lib="Human Fetal Pancreas 1B"  
 /rissue\_type="Fetal Pancreas (4 Pooled Donors, 18 - 20  
 weeks, StrataGene #738023)"  
 /dev\_stage="Fetal Pancreas"  
 /note="Vector: pBluescript SK(-); Site 1: NotI; Site 2:  
 XhoI; cDNA made by oligo-dT priming. Size-selected on  
 agarose gel. Average insert size ~1kb. 5' XhoI site was  
 destroyed after directional cloning. Amplified once.  
 Contact information: Hiroshi Inoue, MD, Metabolism Div.  
 (Alan Permut Lab), Washington University School of  
 Medicine, Box 8127, 660 South Euclid Ave., St. Louis, MO  
 63110 E-mail: hinoue@imgate.wustl.edu Tel: 314-362-1916,  
 Fax: 314-747-2692."

BASE COUNT 133 a 163 c 132 g 115 t

ORIGIN

Query Match 35.2%; Score 492; DB 13; Length 543;  
 Best local similarity 99.8%; Pred. No. 0;  
 Matches 542; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 814 TCCCACTGGGGCTGTGCTCCCTCTGAGACGGTTCGTGGACAGCCCCCATCATCTGTGT 873  
 DB 543 TCCCACTGGGGCTGTGCTCCCTCTGAGACGGTTCGTGGACAGCCCCCATCATCTGTGT 484

QY 874 TCAATGTGTGAAGATGCTAAAGCCCTGTGTGTCTGTGTGTGTGTGTGTGTGTGTGTGTGT 933  
 DB 483 TCAATGTGTGAAGATGCTAAAGCCCTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 424

QY 934 CGGTGGGGCTGCTGGGGACAAATCCATGTCGAGCTTCTCTGAGCTTAGCTTGACA 993  
 DB 423 CGGTGGGGCTGCTGGGGACAAATCCATGTCGAGCTTCTCTGAGCTTAGCTTGACA 364

QY 994 GGAGACTTGGCGGGGATGCTCCAGAGATGTGGTATTTCTGACTTGGGGAGGCTATCTTC 1053  
 DB 363 GGAGACTTGGCGGGGATGCTCCAGAGATGTGGTATTTCTGACTTGGGGAGGCTATCTTC 304

QY 1054 TGACCTCCCGACAGGAGCACTCCAGGCGAGCCAGGGGCTAGGGGCGAGAGGTGACAC 1113  
 DB 303 TGACCTCCCGACAGGAGCACTCCAGGCGAGCCAGGGGCTAGGGGCGAGAGGTGACAC 244

QY 1114 CTCACATGAGCCAAAGACTGGGGTCAAGGAGCGGTGTGTTGAGCCAGGACTTGGGGC 1173  
 DB 243 CTCACATGAGCCAAAGACTGGGGTCAAGGAGCGGTGTGTTGAGCCAGGACTTGGGGC 184

QY 1174 GGGGGTGGGGCGGGGCTTTTCTGCTCATTTTGTCTTCAATGAAAGCTCAAGAGCA 1233  
 DB 183 GGGGGTGGGGCGGGGCTTTTCTGCTCATTTTGTCTTCAATGAAAGCTCAAGAGCA 124

QY 1234 AAACGAGCTTCCCTCTCGAGTTGATATCCGAATCTTTGTACTTCTGTG 1293  
 DB 123 AAACGAGCTTCCCTCTCGAGTTGATATCCGAATCTTTGTACTTCTGTG 64

QY 1294 GTTAATGTTTATTTTGTAAATAAATAAATAAATTAGTTAATAAATGATGTTTCA 1353  
 DB 63 GTTAATGTTTATTTTGTAAATAAATAAATAAATTAGTTAATAAATGATGTTTCA 4

QY 1354 AGC 1356  
 DB 3 AGC 1

RESULT 9  
 BE677292/c 504 bp mRNA linear EST 08-SEP-2000  
 LOCUS 7d83b01.x1 Lupski dorsal root ganglion Homo sapiens cDNA clone  
 DEFINITION IMAGE:3279529.3', similar to, contains MER11, c3 MSRI repetitive  
 element; mRNA sequence.  
 ACCESSION BE677292.1 GI:10037833  
 VERSION BE677292.1  
 KEYWORDS EST.  
 SOURCE human.  
 ORGANISM Homo sapiens  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.  
 REFERENCE 1 (bases 1 to 504)  
 AUTHORS NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.  
 TITLE National Cancer Institute, Cancer Genome Anatomy Project (CGAP),  
 Tumor Gene Index  
 JOURNAL Unpublished (1997)  
 COMMENT Contact: Robert Strausberg, Ph.D.  
 Email: cgapbs-r@mail.nih.gov  
 Tissue Procurement: Dr. James Lupski  
 cDNA Library Preparation: Lupski Laboratory  
 DNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)  
 DNA Sequencing by: Washington University Genome Sequencing Center  
 Clone distribution: NCI-CGAP clone distribution information can be  
 found through the I.M.A.G.E. Consortium/LLNL at:  
 info@image.llnl.gov  
 Seg primer: -40UP from Gibco  
 High quality sequence stop: 409.  
 Location/Qualifiers  
 1. 504  
 /organism="Homo sapiens"  
 /db\_xref="taxon:9606"  
 /clone\_image="IMAGE:3279529"  
 /clone\_lib="Lupski\_dorsal\_root\_ganglion"  
 /sex="male"  
 /rissue\_type="dorsal root ganglia"  
 /dev\_stage="adult, 36 yr"  
 /lab\_host="DH10B"  
 /note="Vector: pCMV-SPORT6 (Life Technologies); Site 1:

NotI, Site 2: SalI; cDNA made by oligo-dT priming. Directionally cloned using the following adaptors: 5'-TGACCCAGGCGCCG-3' and 5'-GACTAGTCTAGTACGAGGCGCCCT(15)-3'. Size selected > 1 kb for average insert length 1.7 kb. This is a primary library, non-amplified. Library constructed by Life Technologies and donated by J. Lupski, M.D./Ph.D. (Baylor College of Medicine) and is available through Life Technologies."

## BASE COUNT

124 a 147 c 115 g 118 t

## ORIGIN

Query Match 34.8%; Score 486; DB 10; Length 504;  
Best Local Similarity 100.0%; Pred. No. 0;  
Matches 486; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 858 AGCCCATCATCTGTTTCAATAGTGTGAGATAGTAAAGCCCTGCTGCTGCTG 917  
DB 504 AGCCCATCATCTGTTTCAATAGTGTGAGATAGTAAAGCCCTGCTGCTGCTG 445  
QY 918 CACATGCCACAGAGGCGGTGCGGTGCGGTGCGGTGCGGTGCGGTGCGGTG 977  
DB 444 CACATGCCACAGAGGCGGTGCGGTGCGGTGCGGTGCGGTGCGGTGCGGTG 385  
QY 978 AGCTTAGTGTGACAGAGGAGATTTGGCGGGGATGCTCCAGAGATGGGTGATTTCTGAC 1037  
DB 384 AGCTTAGTGTGACAGAGGAGATTTGGCGGGGATGCTCCAGAGATGGGTGATTTCTGAC 325  
QY 1038 CTGGGAGAGGCTATCTGACCTCCCGACAGGGGACATCCCGAGCCAGCCAGGGGTGAG 1097  
DB 324 CTGGGAGAGGCTATCTGACCTCCCGACAGGGGACATCCCGAGCCAGCCAGGGGTGAG 265  
QY 1098 GGGGAGAGGTGACACCTCAGATGAGCAAGATGCGGGGTGAGGGAGAGAGTGTGTTG 1157  
DB 264 GGGGAGAGGTGACACCTCAGATGAGCAAGATGCGGGGTGAGGGAGAGAGTGTGTTG 205  
QY 1158 AGCCAGAGATCTGGGGCGGGGCGGGGCGGGGCGGGGCGGGGCGGGGCGGGG 1217  
DB 204 AGCCAGAGATCTGGGGCGGGGCGGGGCGGGGCGGGGCGGGGCGGGGCGGGG 145  
QY 1218 AGCTCAAGAGAGCAAAACAGAGCTTTCCCTCTCCAGATTTGAATATCCAGATCT 1277  
DB 144 AGCTCAAGAGAGCAAAACAGAGCTTTCCCTCTCCAGATTTGAATATCCAGATCT 85  
QY 1278 TTGTGACTCTTGTGTTGTTAAATTTTATTTTGTAAAAATTAATTAATTAAGTTAA 1337  
DB 84 TTGTGACTCTTGTGTTGTTAAATTTTATTTTGTAAAAATTAATTAATTAAGTTAA 25  
QY 1338 TAAAT 1343  
DB 24 TAAAT 19

RESULT 10  
EM726293 543 bp mRNA linear EST 01-MAR-2002  
LOCUS UI-E-BJ0-aih-d-23-0-UI-r1 UI-E-BJ0 Homo sapiens cDNA clone  
DEFINITION BM726293  
ACCESSION BM726293  
VERSION BM726293.1 GI:19047626  
KEYWORDS EST.  
SOURCE human.  
ORGANISM Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.  
REFERENCE 1 (bases 1 to 543)  
AUTHORS Bonaldo, M.F., Lennon, G. and Soares, M.B.  
TITLE Normalization and subtraction: two approaches to facilitate gene  
discovery  
JOURNAL Genome Res. 6 (9), 791-806 (1996)  
MEDLINE 97044477  
COMMENT Program for Rat Gene Discovery and Mapping

## FEATURES

## source

University of Iowa  
451 Eckstein Medical Research Building Iowa City, IA 52242, USA  
Tel: 319 335 8250  
Fax: 319 335 9565  
Email: mesores@blue.weeg.uiowa.edu  
Tissue Procurement: Dr. Gregg Hageman  
cDNA Library preparation: Dr. M. Bento Soares, University of Iowa  
cDNA Library Arrayed by: Dr. M. Bento Soares, University of Iowa  
DNA Sequencing by: Dr. M. Bento Soares, University of Iowa  
Clone Distribution: Researchers may obtain clones from Research  
Genetics (www.resgen.com).  
Seq primer: M13 Reverse.

## Location/Qualifiers

1. 543  
/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
/clone="UI-E-BJ0-aih-d-23-0-UI"  
/clone\_1ib="UI-E-BJ0"  
/tissue\_type="fetal eyes, lens, eye anterior segment,  
optic nerve, retina, Retina Foveal and Macular, RPE and  
Choroid"  
/dev\_stage="fetal and adult"  
/lab\_host="pH10B (Life Technologies) (T1 phage resistant)"  
/note="Organ: eye; Vector: pT73-Pac (Pharmacia) with a  
modified polylinker; Site 1: Ecor I; Site 2: Not I;  
UI-E-BJ0 is a subtracted cDNA library constructed  
according to Bonaldo, Lennon and Soares, Genome Research,  
6:791-806, 1996. First strand cDNA synthesis was primed  
with an oligo-dT primer containing a Not I site. Double  
stranded cDNA was ligated to an Ecor I adaptor, digested  
with Not I, and cloned directionally into the pT73-Pac  
vector. The oligonucleotide used to prime the synthesis of  
first-strand cDNA contains a library tag sequence that is  
located between the Not I site and the (dT)18 tail. The  
sequence tags for this library are: fetal eyes, AGAATCAGA  
; lens, CGATTAGCGA; eye anterior segment, AATGCCGAT;  
optic nerve, CCATTAAGTG; retina, CCGCG; Retina Foveal and  
Macular, GTCC; RPE and Choroid, ACCCTA. This library was  
created for the program, Gene Discovery in the Visual  
System, supported by National Eye Institute (NEI)."

## BASE COUNT

120 a 155 c 150 g 118 t

## ORIGIN

Query Match 34.8%; Score 486; DB 14; Length 543;  
Best Local Similarity 100.0%; Pred. No. 0;  
Matches 486; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 442 CAGGTGGGAGCTGCTGTTTCAAGATGAGTGAAGCCCAAGTGTCTCTATCCCA 501  
DB 58 CAGGTGGGAGCTGCTGTTTCAAGATGAGTGAAGCCCAAGTGTCTCTATCCCA 117  
QY 502 AGACCAATATGTGAAGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 561  
DB 118 AGACCAATATGTGAAGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 177  
QY 562 ATCTCAGAGCTGGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 621  
DB 178 ATCTCAGAGCTGGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 237  
QY 622 ACCTGGGAACCTCTGCAAGAAAGCTAATCTCCAGACAGAGATGTATCCAGACAA 681  
DB 238 ACCTGGGAACCTCTGCAAGAAAGCTAATCTCCAGACAGAGATGTATCCAGACAA 297  
QY 682 CGTGCAATATGTGAAGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 741  
DB 298 CGTGCAATATGTGAAGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 357  
QY 742 TCCTAGTCCAGAGATATGAGGGGTATGATGCTCTTCAACCTGTGGCTGTAGCAAG 801  
DB 358 TCCTAGTCCAGAGATATGAGGGGTATGATGCTCTTCAACCTGTGGCTGTAGCAAG 417  
QY 802 CTGAGGCTATCTTCCCACTGGGGGCTGTGCTCTCTGAGGAGGTTCCGTGGGAGGC 861



Db 418 CTCAGCTGATCTCCCTCCACTGGGGGCTGTGCCCCCTCCCTGGGACGGTTCCGTGGGACGCC 477  
QY 862 CCATCACTGTTCATAGTGTAGATGTAGTAAAGCCCTGCTGCTGTCTGCACA 921  
Db 478 CCATCACTGTTCATAGTGTAGATGTAGTAAAGCCCTGCTGCTGTCTGCACA 537  
QY 922 TGGCAC 927  
Db 538 TGGCAC 543

RESULT 11  
A1453409 492 bp mRNA linear EST 13-APR-1999  
A1453409/c t37h12.x1 NCI\_CGAP\_Pan1 Homo sapiens cDNA clone IMAGE:2143751 3',  
DEFINITION mRNA sequence.  
ACCESSION A1453409.1 GI:4281679  
VERSION A1453409  
KEYWORDS EST.  
SOURCE human.  
ORGANISM Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Hominiidae; Homo.  
REFERENCE 1 (bases 1 to 492)  
NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.  
AUTHORS National Cancer Institute, Cancer Genome Anatomy Project (CGAP),  
TITLE Tumor Gene Index  
JOURNAL Unpublished (1997)  
COMMENT Contact: Robert Strausberg, Ph.D.  
Email: cgapbs-remail.nih.gov  
Life Technologies catalog #: 11548-013  
DNA Sequencing by: Washington University Genome Sequencing Center  
Clone distribution: NCI-CGAP clone distribution information can be  
found through the I.M.A.G.E. Consortium/LLNL at:  
www.bio.llnl.gov/bbrp/image/image.html  
Insert Length: 1208 Std Error: 0.00  
Seq primer: -40UP from Gibco  
High quality sequence stop: 438.  
Location/Qualifiers  
1..492  
/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
/clone\_1ib="IMAGE:2143751"  
/clone\_1ib="NCI CGAP Pan1"  
/tissue\_type="adenocarcinoma"  
/lab\_host="DH10B"  
/note="Organ: pancreas; Vector: pCMV-SPORT6; Site 1: SalI;  
Site 2: NotI; Cloned unidirectionally. Primer: Oligo dt.  
Average insert size 1.72-kb. Life Technologies catalog #:  
11548-013"

BASE COUNT 123 a 146 c 112 g 111 t

Query Match 34.2%; Score 478; DB 9; Length 492;  
Best Local Similarity 100.0%; Pred. No. 0;  
Matches 478; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 866 CACTGTGTCATAGTGTAGATGTAGTAAAGCCCTGCTGCTGTCTGCACA 925  
Db 492 CACTGTGTCATAGTGTAGATGTAGTAAAGCCCTGCTGCTGTCTGCACA 433  
QY 926 ACAGCAGGCGGTGGGGCTGCGTGGGACATCCATCGTGAAGTCTCTCAGCTTAG 985  
Db 432 ACAGCAGGCGGTGGGGCTGCGTGGGACATCCATCGTGAAGTCTCTCAGCTTAG 373  
QY 986 TCTGACAGAGACTTGGCGGGGATGCTCCAGAGTGTGGTGATTTCTGTACTGGGAG 1045  
Db 372 TCTGACAGAGACTTGGCGGGGATGCTCCAGAGTGTGGTGATTTCTGTACTGGGAG 313  
QY 1046 GCATCTGACCTCCGACAGGGGACATCCCGACGCCAGGGGTCCAGGGGAGAG 1105  
Db 312 GCATCTGACCTCCGACAGGGGACATCCCGACGCCAGGGGTCCAGGGGAGAG 253

QY 1106 GTGACACCTCAGCATGAGCCAGAGCTGGGGCTACAGGAGCAGGTGTGTTAGCCAGCA 1165  
Db 252 GTGACACCTCAGCATGAGCCAGAGCTGGGGCTACAGGAGCAGGTGTGTTAGCCAGCA 193  
QY 1166 CTTGGGGGGGGGGTGGGGGCTTTTCTGCTCATTTGCTTTCAATGAAGCTCA 1225  
Db 192 CTTGGGGGGGGGGTGGGGGCTTTTCTGCTCATTTGCTTTCAATGAAGCTCA 133  
QY 1226 AGCAGCCAAAACAGGCTTTCCCTCCCTCAGTTTGAATATCCAGATCTTTGTACT 1285  
Db 132 AGCAGCCAAAACAGGCTTTCCCTCCCTCAGTTTGAATATCCAGATCTTTGTACT 73  
QY 1286 TCTTGTGTTAAATGTTTATTTTGAATAAATAAATAAATAGTTAAATAAT 1343  
Db 72 TCTTGTGTTAAATGTTTATTTTGAATAAATAAATAAATAAATAGTTAAATAAT 15

RESULT 12  
BQ012938 485 bp mRNA linear EST 26-MAR-2002  
LOCUS BQ012938/c  
DEFINITION UI-1-BC1P-ava-g-11-0-UI.s1 NCI CGAP Pl3 Homo sapiens cDNA clone  
UI-1-BC1P-ava-g-11-0-UI 3', mRNA sequence.  
ACCESSION BQ012938  
VERSION BQ012938.1 GI:19737839  
KEYWORDS EST.  
SOURCE human.  
ORGANISM Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Hominiidae; Homo.  
REFERENCE 1 (bases 1 to 485)  
NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.  
AUTHORS National Cancer Institute, Cancer Genome Anatomy Project (CGAP),  
TITLE Tumor Gene Index  
JOURNAL Unpublished (1997)  
COMMENT Contact: Robert Strausberg, Ph.D.  
Email: cgapbs-remail.nih.gov  
Tissue Procurement: Dr. Steven Brown  
cDNA library preparation: Dr. M. Bento Soares, University of Iowa  
cDNA library Arrayed by: Dr. M. Bento Soares, University of Iowa  
DNA Sequencing by: Dr. M. Bento Soares, University of Iowa  
Clone Distribution: Clone distribution information can be obtained  
from Dr. M. Bento Soares, bento-soares@uiowa.edu  
The following repetitive elements were found in this cDNA  
sequence: 29-77, >AT rich#low\_complexity  
Seq primer: M13 FORWARD  
PolyA=yes.  
Location/Qualifiers  
1..485  
/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
/clone\_1ib="UI-1-BC1P-ava-g-11-0-UI"  
/clone\_1ib="NCI CGAP Pl3"  
/tissue\_type="Placenta"  
/dev\_stage="8-9 weeks"  
/lab\_host="DH10B (Life Technologies)"  
/note="Organ: Placenta; Vector: pRTT3-Pac (Pharmacia) with  
a modified polylinker; Site 1: EcoR I; Site 2: Not II;  
NCI CGAP Pl3 is a subcloned cDNA library constructed  
according to Bonaldo, Lennon and Soares, Genome Research,  
6:791-806, 1996. First strand cDNA synthesis was primed  
with an oligo-dT primer containing a Not I site. Double  
stranded cDNA was ligated to an EcoR I adaptor, digested  
with Not I, and cloned directionally into pRTT3-Pac  
vector. The oligonucleotide used to prime the synthesis of  
first strand cDNA contains a library tag sequence that is  
located between the Not I site and the (dT)18 tail. The  
sequence tags for this library are GA, ACGAA. For  
additional information, contact: Bento Soares,  
bento-soares@uiowa.edu  
TAG\_LIB=UI-1-BC1P  
TAG\_TISSUE=Placenta human 8 week  
TAG\_SEQ=GA"

BASE COUNT 116 a 141 c 108 g 120 t



This clone is available royalty-free through LNL; contact the IMAGE Consortium (info@image.llnl.gov) for further information.  
 Seq primer: -40UP from Gibco  
 High quality sequence stop: 502.  
 Location/Qualifiers

## FEATURES

source

```

1. 522
  /organism="Homo sapiens"
  /db_xref="taxon:9606"
  /clone="IMAGE:2934556"
  /clone_1ib="Soares_NFL_T_GBC_S1"
  /lab_host="DH10B"
  /note="Organ: pooled; Vector: pRTT3D-Pac (Pharmacia) with
  a modified polylinker; Site 1: Not I; Site 2: Eco RI;
  Equal amounts of plasmid DNA from three normalized
  libraries (fetal lung NbHL19W, testis NHT, and B-cell
  NCI CGAP GC81) were mixed, and as circles were made in
  vitro. Following HAP purification, this DNA was used as
  tracer in a subtractive hybridization reaction. The driver
  was PCR-amplified cDNAs from pools of 5,000 clones made
  from the same 3 libraries. The pools consisted of
  I.M.A.G.E. clones 297480-302087, 682632-687239,
  726408-728711, and 729096-731399. Subtraction by Bento
  Soares and M. Fatima Bonaldo."
  BASE COUNT      129 a      156 c      121 g      115 t      1 others
  ORIGIN

```

Query Match 33.4%; Score 467; DB 10; Length 522;

Best Local Similarity 99.8%; Pred. No. 0;

Matches 517; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

840 TGGGACGGTCCGTGGGACAGCCCATCATCTGTTCAATAGTGAAGTGTAAAG 899

518 TGGGACGGTCCGTGGGACAGCCCATCATCTGTTCAATAGTGAAGTGTAAAG 459

900 CCCCTGCTGCTGCTGCTGCACATGCCACAGACGGGCTGCGTGGGACATCC 959

458 CCCCTGCTGCTGCTGCTGCACATGCCACAGACGGGCTGCGTGGGACATCC 399

960 ATCGTAGAGTGTCTCTCAGACTTAGGTCTGGAAGAGAGCTTGGCGGGAGTCTCCAG 1019

398 ATGTGAGAGTGTCTCTCAGACTTAGGTCTGGAAGAGAGCTTGGCGGGAGTCTCCAG 339

1020 ATGTGAGAGTGTCTCTCAGACTTAGGTCTGGAAGAGAGCTTGGCGGGAGTCTCCAG 1079

338 ATGTGAGAGTGTCTCTCAGACTTAGGTCTGGAAGAGAGCTTGGCGGGAGTCTCCAG 279

1080 GGGCAGCCAGGGGTGAGGGGCGAGAGTGTGACACTTCAGCATAGCCAAAGCTGGGCTCA 1139

278 GGGCAGCCAGGGGTGAGGGGCGAGAGTGTGACACTTCAGCATAGCCAAAGCTGGGCTCA 219

1140 GGGAGCAGGTGTGTTGAGCCAGACCTGGGGCGGGGGTGGGGCGGGGCTTTCGCC 1199

218 GGGAGCAGGTGTGTTGAGCCAGACCTGGGGCGGGGGTGGGGCGGGGCTTTCGCC 159

1200 TCATTGCTTTCAATGAAGGCTCAAAAGCAGCCAAACAGGCTTCCCTTCCTCGAG 1259

158 TCATTGCTTTCAATGAAGGCTCAAAAGCAGCCAAACAGGCTTCCCTTCCTCGAG 99

1260 TTTGAATATTCAGAAATCTTTGTTACTCTTGTGTTGTTAAATGTTTATTTTGTAAAAA 1319

98 TTTGAATATTCAGAAATCTTTGTTACTCTTGTGTTGTTAAATGTTTATTTTGTAAAAA 39

1320 TAAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTA 1357

38 TAAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTA 1

Db

RESULT 15

BM693225

LOCUS

DEFINITION

UI-E-CK1-afm-c-23-0-UI.r1 UI-E-CK1 Homo sapiens cDNA clone

ACCESSION

## FEATURES

source

```

1. 465
  /organism="Homo sapiens"
  /db_xref="taxon:9606"
  /clone="UI-E-CK1-afm-c-23-0-UI"
  /clone_1ib="UI-E-CK1"
  /tissue_type="Retina Foveal and Macular"
  /dev_stage="adult"
  /lab_host="DH10B (Life Technologies) (T1 phage resistant)"
  /note="Organ: eye; Vector: pRTT3-Pac (Pharmacia) with a
  modified polylinker; Site 1: EcoR I; Site 2: Not I;
  UI-E-CK1 is a normalized cDNA library containing the
  following tissue(s): Retina Foveal and Macular. The
  library was constructed according to Bonaldo, Lennon and
  Soares, Genome Research, 6:791-806, 1996. First strand
  cDNA synthesis was primed with an oligo-dT primer
  containing a Not I site. Double stranded cDNA was ligated
  to an EcoR I adaptor, digested with Not I, and cloned
  directionally into pRTT3-Pac vector. The oligonucleotide
  used to prime the synthesis of first-strand cDNA contains
  a library tag sequence that is located between the Not I
  site and the GTC. This library was created for the program,
  Gene Discovery in the Visual System, supported by National
  Eye Institute (NEI)."
  BASE COUNT      98 a      129 c      137 g      101 t
  ORIGIN

```

Query Match 33.3%; Score 465; DB 14; Length 465;

Best Local Similarity 100.0%; Pred. No. 0;

Matches 465; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

513 GTGAAGGCTGTGGCTGACCTATCTGAGGGCTGGGCTGACAGCTGACTATCTTCAGAG 572

1 GTGAAGGCTGTGGCTGACCTATCTGAGGGCTGGGCTGACAGCTGACTATCTTCAGAG 60

573 CTGGGCTTCCCTGTGAGGAGTGAATCTTGCATCTGGCAGACATGTCATCTGGAAC 632

61 CTGGGCTTCCCTGTGAGGAGTGAATCTTGCATCTGGCAGACATGTCATCTGGAAC 120

633 CTGGCAGCAAGCTTACATCCAGACAGATGTGACAGCAAGCAAGCTGCAATAT 692

121 CTGGCAGCAAGCTTACATCCAGACAGATGTGACAGCAAGCAAGCTGCAATAT 180

693 GCCAAATGTTAAATGTGAGTTTACACAGCTATGAGGAGTGTGCTGCTCTAGTCAG 752

181 GCCAAATGTTAAATGTGAGTTTACACAGCTATGAGGAGTGTGCTGCTCTAGTCAG 240

Db

VERSION

BM693225.1 GI:19006483

EST.

KEYWORDS

SOURCE

ORGANISM

Homo sapiens

human

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

Mammalia; Eutheria; Primates; Catarrhini; Hominiidae; Homo.

1 (bases 1 to 465)

Bonaldo, M.F., Lennon, G. and Soares, M.B.

Normalization and subtraction: two approaches to facilitate gene

discovery

Genome Res. 6 (9), 791-806 (1996)

97044477

CONTACT: Soares, MB

Program for Rat Gene Discovery and Mapping

University of Iowa

451 Eckstein Medical Research Building Iowa City, IA 52242, USA

Tel: 319 335 8250

Fax: 319 335 9565

Email: msoares@blue.weeg.uiowa.edu

Tissue Procurement: Dr. Gregg Hageman

cDNA Library preparation by: Dr. M. Bento Soares, University of Iowa

cDNA Library Arrayed by: Dr. M. Bento Soares, University of Iowa

DNA Sequencing by: Dr. M. Bento Soares, University of Iowa

Clone Distribution: Researchers may obtain clones from Research

Genetics (www.resgen.com).

Seq primer: M13 Reverse.

Location/Qualifiers

```

QY 753 GAATCATGGGGGATAGTACTGCTCTCTCAACCTGTGGGCTGTAGCAAGCTCAGGCTAGT 812
Db 241 GAATCATGGGGGATAGTACTGCTCTCTCAACCTGTGGGCTGTAGCAAGCTCAGGCTAGT 300
QY 813 CTCCCACTGGGGGCTGTGCTCCCTCCCTGGAGGGTTCCGTGGGAGCCCATCCTGNG 872
Db 301 CTCCCACTGGGGGCTGTGCTCCCTCCCTGGAGGGTTCCGTGGGAGCCCATCCTGNG 360
QY 873 TTCAATAGTGTGAGATGTAGTAAAGCCCTGCTGCTGCTGCACATGCCACAGCAG 932
Db 361 TTCAATAGTGTGAGATGTAGTAAAGCCCTGCTGCTGCTGCACATGCCACAGCAG 420
QY 933 GCGGTGGGGGCTGCGTGGGGACAATCCATCGTGAAGTTCTCTC 977
Db 421 GCGGTGGGGGCTGCGTGGGGACAATCCATCGTGAAGTTCTCTC 465

```

Search completed: June 29, 2003, 06:37:13  
 Job time : 1522 secs

GenCore version 5.1.6  
Copyright (c) 1993 - 2003 Compugen Ltd.

OM nucleic - protein search, using frame\_plus\_n2p model

Run on: June 24, 2003, 13:41:02 ; Search time 82.5 seconds  
(without alignments)  
4512.754 Million cell updates/sec

Title: US-09-989-919-15

Perfect score: 2527  
Sequence: 1 ggtgctgcacctctaccgga.....aaaaaaaaaaaaagcgcgtc 1397

Scoring table: BLOSUM62  
Xgapop 10.0, Xgapext 0.5  
Ygapop 10.0, Ygapext 0.5  
Fgapop 6.0, Fgapext 7.0  
Delop 6.0, Delext 7.0

Searched: 908470 seqs, 133250620 residues

Total number of hits satisfying chosen parameters: 1816940

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Command line parameters:  
-MODE=frame+ n2p.model -DEV=xlp  
-O=/cgcn2.1/USPTO\_spoop/US09989919/runat\_24062003\_102045\_19166/app\_query.fasta\_1.1543  
-DB=A\_Genseq\_101002 -QMT=faetan -SUFFIX=tag -MINMATCH=0.1 -LOOPEXT=0  
-LOOPEXT=0 -UNITS=bits -START=1 -END=1 -MATRIX=blosum62 -TRANS=human40.cdi  
-LIST=45 -DOCALLIGN=200 -THR SCORE=pct -THR MAX=100 -THR MIN=0 -ALIGN=15  
-MODE=LOCAL -OUTFMT=ptc -NORM=ext -HEAPSIZE=500 -MINLEN=0 -MAXLEN=2000000000  
-USER=US09989919 @CGN 1.114 @runat\_24062003\_102045\_19166 -NCPU=6 -ICPU=3  
-NO MAP -LARGEOUTERY -NEG\_SCORES=0 -WAIT -DSPBLOCK=100 -LONGLOC  
-DEV TIMEOUT=120 -MARN TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOP=6  
-FGAPEXT=7 -YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database: A\_Genseq\_101002:  
1: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1980.DAT:  
2: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1981.DAT:  
3: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1982.DAT:  
4: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1983.DAT:  
5: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1984.DAT:  
6: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1985.DAT:  
7: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1986.DAT:  
8: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1987.DAT:  
9: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1988.DAT:  
10: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1989.DAT:  
11: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1990.DAT:  
12: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1991.DAT:  
13: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1992.DAT:  
14: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1993.DAT:  
15: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1994.DAT:  
16: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1995.DAT:  
17: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1996.DAT:  
18: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1997.DAT:  
19: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1998.DAT:  
20: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1999.DAT:  
21: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA2000.DAT:  
22: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA2001.DAT:  
23: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA2002.DAT:

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	519.5	20.6	540	22 AAB65697	Novel protein kina
2	431	17.1	365	22 AAB59576	Novel protein kina
3	314	12.4	482	22 AAB514651	Novel human diagno
4	168	6.6	63	22 AAM23513	Human EST encoded
5	156.5	6.2	1093	14 AAR41001	Human myocytin dya
6	144.5	5.6	1040	22 ABG14734	Novel human diagno
7	139.5	5.4	1435	22 ABB11808	Human BCL9 homolog
8	134.5	5.3	738	23 AAU97643	Human Hsa011374 pr
9	125	4.9	384	22 AAU32775	Novel human secret
10	124	4.8	1339	15 AAR45945	Glutamic acid rece
11	123.5	4.8	505	22 ABG03717	Novel human diagno
12	123.5	4.8	1212	20 AAB75703	Human N-methyl-D-a
13	122.5	4.8	647	22 AAB74205	Brassica napus PER
14	122.5	4.8	900	22 ABG03533	Novel human diagno
15	122	4.8	343	21 AAY44789	Corn Adenylsulph
16	121.5	4.7	522	22 AAU17573	Novel signal trans
17	121.5	4.7	524	22 AAU17196	Novel signal trans
18	121.5	4.7	1400	21 AAB40638	Human OREF ORP402
19	120.5	4.7	871	22 AAM78479	Human protein SEQ
20	120.5	4.7	871	22 AAM79463	Human protein SEQ
21	120	4.7	684	22 ABG14126	Novel human diagno
22	120	4.7	815	22 ABB70499	Drosophila melano
23	120	4.7	1061	20 AAM87504	Human N-methyl-D-a
24	120	4.7	1964	20 AAM95557	Mus musculus notch
25	119.5	4.7	626	22 AAM40777	Human polypeptide
26	119.5	4.7	806	22 ABG07472	Novel human diagno
27	119.5	4.7	806	22 ABG15589	Novel human diagno
28	119.5	4.7	855	22 AAU02936	Angiotensin conver
29	119.5	4.7	1214	15 AAR66065	Human NMDAR2 recep
30	119.5	4.7	1219	15 AAR66063	Human NMDAR2 recep
31	119.5	4.7	1231	15 AAR66062	Human NMDAR2 recep
32	119.5	4.7	1336	15 AAR66037	Human N-methyl-D-a
33	119.5	4.7	1339	15 AAR66064	Human NMDAR2 recep
34	119.5	4.7	1244	15 AAR66061	Human NMDAR2 recep
35	119	4.6	749	22 ABB62915	Drosophila melano
36	119	4.6	897	22 AAU00387	Cytokine receptor
37	119	4.6	897	22 AAU00388	Cytokine receptor
38	119	4.6	1063	16 AAR79049	Infectious rubella
39	118.5	4.6	542	22 ABB11016	Human secreted pro
40	118.5	4.6	1619	23 AAU77498	Human lipid metabo
41	118	4.6	897	13 AAR20982	Sequence of beta-c
42	118	4.6	897	22 AAU00374	Cytokine receptor
43	118	4.6	897	22 AAU00385	Cytokine receptor
44	118	4.7	961	23 AAU80188	Human TSP1 domain
45	118	4.6	1061	20 AAM87504	Human N-methyl-D-a

ALIGNMENTS

RESULT 1  
AAB65697  
ID AAB65697 standard, Protein: 540 AA.  
AC AAB65697;  
XX  
DT 27-MAR-2001 (first entry)  
XX  
DE Novel protein kinase, SEQ ID NO: 225.  
XX  
KW Human; mouse; protein kinase; antiarthritic; antisclerotic; osteoparhnic;  
KW immunosuppressive; cardiac; renal; antiinflammatory; antiaslamic;  
KW dermatological; antidiabetic; antifertility; gene therapy; vaccina;  
KW immune disorder; cardiovascular disease; neurodegenerative disease;  
KW cancer; autoimmune disorder; stroke; inflammatory bowel disease;  
KW inflammatory pelvic disease; multiple sclerosis; psoriasis.  
XX  
OS Homo sapiens.  
XX

PN MO200073469-A2.  
 XX 07-DEC-2000.  
 PD 26-MAY-2000; 2000MO-US14842.  
 PF 28-MAY-1999; 99US-0136503.  
 XX (SUGEN-) SUGEN INC.  
 PA Plowman GD, Martinez R, Whyte D, Sudersanam S;  
 PI WPI: 2001-032161/04.  
 DR N-PSDB; AAF44725.  
 XX Nucleic acids encoding kinase polypeptides, useful for diagnosing and  
 PT treating immune-related diseases and disorders, cardiovascular disease,  
 PT neurodegenerative diseases and/or cancers -  
 XX Claim 10; Fig 1; 310pp; English.  
 XX The present sequence is a novel protein kinase. The novel protein kinases  
 CC and the nucleic acids that encode them may be used in the treatment and  
 CC diagnosis of diseases associated with inappropriate kinase expression  
 CC such as immune-related diseases and/or cancers. The nucleic acids and  
 CC neurodegenerative diseases and/or cancers. The nucleic acids and  
 CC complementary sequences may also be used as DNA probes in diagnostic  
 CC assays. The kinase polypeptides may be used as antigens in the production  
 CC of antibodies of kinase expression and activity. Anti-kinase antibodies  
 CC and kinase antagonists may also be used to down regulate kinase  
 CC expression and activity. Diseases related to kinase expression and  
 CC activity include rheumatoid arthritis, atherosclerosis, autoimmune  
 CC disorders, complications of organ transplantation, myocardial infarction,  
 CC immune disorders, cardiomyopathies, strokes, renal failure,  
 CC oxidative-stress related disorders, chronic inflammatory bowel disease,  
 CC chronic inflammatory pelvic disease, multiple sclerosis, asthma,  
 CC osteoarthritis, psoriasis, rhinitis, autoimmunity, diabetes, cancers and  
 CC reproductive disorders.  
 CC Sequence 540 AA;  
 SQ  
 Alignment Scores:  
 Pred. No.: 6e-40 Length: 540  
 Score: 519.50 Matches: 105  
 Percent Similarity: 59.66% Conservative: 0  
 Best Local Similarity: 59.66% Mismatches: 1  
 Query Match: 20.56% Indels: 70  
 DB: Caps: 1  
 US-09-989-919-15 (1-1397) x IAB65697 (1-540)  
 QY 2 GTGTGACCTGTATCCGAGGCGGACAGTATCTGCAGAACTCCACGCGAGCAGACAGTACC 61  
 DB 435 ValLeuHisLeuTyrArgSerGlyGlnTyrLeuInLanSerThrAlaSerSerThr 454  
 QY 62 GAGTACCAAGTATATCCACAGACACATATCCCGAGAAACTACCCCTGCGCCATCC 121  
 DB 455 GlnTyrGlnCysIleProAspSerThrIleProInLanAspTyrArgCysTyrProSer 474  
 QY 122 TACACACAGGAGGAGCTCCCTCTTCAAGTGTCACTGAGGCTGAGGCTGTGTGT 181  
 DB 475 TyrHisHisGlySerCysLeuLeuSerValPheHisLeuAlaGlnValAlaValCys 494  
 QY 182 GAGAGCATGCGCCAGTGTGGGCTTTGTGTGTCAACACAGACACACCTGAGCAGGTGAG 241  
 DB 495 GluSerHisAlaGlnCysArgAlaPheValValThrSngInThrTyrThrGlyArg 514  
 QY 242 CCAAGTGGAGAGAGCCCTTCCAGAGGAATGACAGACCTCTCTGAGAGTTGATGATAGT 301  
 DB 515 Gln----- 515  
 QY 302 GATCCCCATCGAAGTCAGAGGGGTGCTGAGGTATGAGAGAGAGATATACGTGTCTT 361

DB 515 ----- 515  
 QY 362 CAAGCAGTCAATATAGGAGAGATGTCTTGTCTCCAGAAAGAAACATCAGCCCTGT 421  
 DB 515 ----- 515  
 QY 422 TACTCTCATCCTTGCCCCCAGAGTCGAGAGTGTCTTTTTCAGACTGTGATGAGCA 481  
 DB 516 -----  
 QY 482 ACTGTCTCTGTATCCCAACAGACCAATATGTGAAAGCCCTGTGC 527  
 DB 525 nValValProAspProAsnLysThrThrTyrValLysAlaSerGly 540  
 RESULT 2  
 AAB59576  
 ID AAB59576 standard; Protein; 365 AA.  
 AC AAB59576;  
 XX 27-MAR-2001 (first entry)  
 DT Novel protein kinase, SEQ ID NO: 226.  
 DE Human; mouse; protein kinase; antiarthritic; antisclerotic; osteopathic;  
 KW immunosuppressive; cardiac; renal; antiinflammatory; antiasthmatic;  
 KW dermatological; antidiabetic; antifertility; gene therapy; vaccine;  
 KW immune disorder; cardiovascular disease; neurodegenerative disease;  
 KW cancer; autoimmune disorder; stroke; inflammatory bowel disease;  
 KW inflammatory pelvic disease; multiple sclerosis; psoriasis.  
 XX Mus musculus.  
 OS  
 XX MO200073469-A2.  
 PN 07-DEC-2000.  
 PD 26-MAY-2000; 2000MO-US14842.  
 XX 28-MAY-1999; 99US-0136503.  
 PR (SUGEN-) SUGEN INC.  
 PA Plowman GD, Martinez R, Whyte D, Sudersanam S;  
 PI WPI: 2001-032161/04.  
 DR N-PSDB; AAF44726.  
 XX Nucleic acids encoding kinase polypeptides, useful for diagnosing and  
 PT treating immune-related diseases and disorders, cardiovascular disease,  
 PT neurodegenerative diseases and/or cancers -  
 XX Claim 10; Fig 1; 310pp; English.  
 XX The present sequence is a novel protein kinase. The novel protein kinases  
 CC and the nucleic acids that encode them may be used in the treatment and  
 CC diagnosis of diseases associated with inappropriate kinase expression  
 CC such as immune-related diseases and/or cancers. The nucleic acids and  
 CC neurodegenerative diseases and/or cancers. The nucleic acids and  
 CC complementary sequences may also be used as DNA probes in diagnostic  
 CC assays. The kinase polypeptides may be used as antigens in the production  
 CC of antibodies of kinase expression and activity. Anti-kinase antibodies  
 CC and kinase antagonists may also be used to down regulate kinase  
 CC expression and activity. Diseases related to kinase expression and  
 CC activity include rheumatoid arthritis, atherosclerosis, autoimmune  
 CC disorders, complications of organ transplantation, myocardial infarction,  
 CC immune disorders, cardiomyopathies, strokes, renal failure,  
 CC oxidative-stress related disorders, chronic inflammatory bowel disease,  
 CC chronic inflammatory pelvic disease, multiple sclerosis, asthma,  
 CC osteoarthritis, psoriasis, rhinitis, autoimmunity, diabetes, cancers and  
 CC reproductive disorders.  
 CC Sequence 365 AA;  
 SQ



```

QY 416 CCCTGTACTCTCACCCTCTGCCGCCAGGTGGGAGCTGGTCTTTTGAAGACTGGATG 475
DB 317 -----TrrpLysMetThrLeuLeuAlaG 324
QY 476 GAGCCAAAGTGGTCCCTGATGCCAACAAGACACATATGTAAGGCTCTGGTACCTA- 534
DB 324 |y|lellelprobserserthrnglyserargpro-----AlaAspLeuA 338
QY 535 -----TGTGAGGGCTGGGCTGACCAAGCTACTATCTTCAGCAGCTGGG 577
DB 338 sptYrPheLeuGlyserSerGlnCysProArgAspGlnArgThrLeuLysGlyHisValS 358
QY 578 CTTCGCTGTGGAGGAGGTACTGCACTGGACACGATGACATGGGAACCCCTGC 637
DB 358 erCys-----AspProThrSerSerIleThrGlyGlyLysLeu 373
QY 638 AGACAAGCTAATCATCCAGACAGACAGATGTGACAGACAAACGTGCATATATCCAA 697
DB 373 ysglnLysLeuGlnSerSerSerThrLeuAlaAsp---ThrLysSerAlaPheLeuGlnA 392
QY 698 ATGTTAAATGTGATTTACACAGCTTACTATGAGGACTGCTGC----- 741
DB 392 sp-----SerTrpThrSerSerLeuProThrLeuGlyAlaSerAlaIleArgT 408
QY 742 -----TCCT 745
DB 408 hrThrSerAlaSer***ThrAspGlnCysProProAlaProGlyProLeuHisValSerP 428
QY 746 AGTCCAGAAATCATGAGGGGTATGATGCTCTCCAAACCTGTGGCTGTAAAGACCTCA 805
DB 428 heserLeuAsn-----SerThrLyrProserArgLeuCysSerSerThrCys 444
QY 806 GGGTACTCTCCCACTGGGGGCTGTGCGCTCCCTCGGAGCGATTCCTGGGAGACCCCAT 865
DB 444 erThrValAspSerPheGlnIleIleProMetProAsn**MetProPhe----- 460
QY 866 CACTGTGTTCAATAGTGTGAATGTAGTAAGCCCTGCTCTGCTGTGCACATGC 924
DB 461 -----SerThrCysGlyAlaPheCysTrpCys 469

RESULT 4
AAM23513
ID AAM23513 standard; Protein; 63 AA.
AC AAM23513;
DT 12-OCT-2001 (first entry)
DE Human EST encoded protein SEQ ID NO: 1038.
XX Human; sheep; pig; cow; fruit fly; yeast; hamster; macaque; horse;
XX tomato; monkey; dog; sea urchin; expressed sequence tag; EST;
XX diagnostics; forensic test; gene mapping; genetic disorder;
XX biodiversity; gene therapy; nutrition.
OS Homo sapiens.
XX
XX WO200154477-A2.
XX
XX 02-AUG-2001.
XX
XX 25-JAN-2001; 2001WO-US02687.
XX
XX 25-JAN-2000; 2000US-0491404.
XX 17-JUL-2000; 2000US-0617746.
XX 03-AUG-2000; 2000US-0631451.
XX 15-SEP-2000; 2000US-0663870.
XX
XX (HYSE-) HYSEQ INC.
XX
XX Tang YT, Liu C, Zhou P, Qian XB, Wang Z, Chen R, Asundi V,
XX Cao Y, Drmanac RA, Zhang J, Werhman T,

```

```

XX MPI: 2001-476164/51.
DR N-PSDB; AAM98172.
XX
PT Isolated polypeptide for treatment of diseases, diagnostics, raising
PT antibodies and research use
PS Claim 20; Page 795; 1275pp; English.
XX
CC The present invention provides the protein and coding sequences of novel
CC proteins from a variety of organisms, including human, dog, cat, horse,
CC cow, pig, hamster, monkey, macaque, yeast, bacteria, fruit fly, sea
CC urchin and tomato. These were derived from expressed sequence tags (ESTs)
CC from the organism of interest. They can be used in diagnostics,
CC forensics, gene mapping, identification of mutations, to assess
CC biodiversity and for nutritional purposes. The present sequence is a
CC protein of the invention.
XX
SQ Sequence 63 AA;
XX
Alignment Scores:
Pred. No.: 3.09e-07 Length: 63
Score: 168.00 Matches: 30
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 6.56% Indels: 0
DB: 22 Gaps: 0

US-09-989-919-15 (1-1397) x AAM23513 (1-63)
QY 90 ATGTGCTGTCTGGGATACACTGTAAGTGGTACTGCTGCTGCGGTGGAGTTCTGCAGA 31
DB 1 MetValLeuSerGlyIleHisTrpLysSerValLeuLeuAlaValGlnPheCysArg 20
QY 30 TACTGCCGCTCGGTATCAGGTGACACACC 1
DB 21 TyrCysProLeuArgTyrArgCysSerThr 30

RESULT 5
AAR41001
ID AAR41001 standard; Protein; 1093 AA.
AC AAR41001;
DT 25-FEB-1994 (first entry)
DE Human myotonic dystrophy gene protein.
XX Abnormality; muscular dystrophy; CHR 19; chromosome 19;
XX protein kinase; polymerase chain reaction; brain.
XX
XX Homo sapiens.
XX
XX Key Location/Qualifiers
XX Region 1..1093
XX FT /note= "encoded by predicted reading frame a.
XX FT x's in the sequence indicate stop codons
XX FT in the reading frame"
XX
XX WO9317104-A.
XX
XX 02-SEP-1993.
XX
XX 19-FEB-1993; 93WO-US01545.
XX
XX 20-FEB-1992; 92US-0839255.
XX
XX (MAST ) MASSACHUSETTS INST TECHNOLOGY.
XX
XX Brook JD, Houseman DE;
XX
XX MPI; 1993-288410/36.

```



PT DNA sequence of myotonic dystrophy gene - used to produce probes  
 PT and identify CHR 19 abnormality and protein kinase responsible  
 XX  
 XX  
 PS Disclosure; Fig 6; 64pp; English.

CC The sequence is that encoded by predicted reading frame a of  
 CC the human myotonic dystrophy (DM) gene. It may be used in the  
 CC identification of individuals affected by DM.

XX  
 XX  
 SQ Sequence 1093 AA;

# Alignment Scores:

Pred. No.:	Length:	Matches:	Conservative:	Mismatches:	Indels:	Gaps:
Score:	1,076-05	1093	127	52	162	202
Percent Similarity:	32.97%					
Best Local Similarity:	23.39%					
Query Match:	6.19%					
DB:	14					28

US-09-989-919-15 (1-1397) x AAR41001 (1-1093)

```

QY 24 GGCAGATATGCG-----AGACTCCAGCGGACAGCAGTACCGAGTACAGTGA 74
DB 605 GlySerThrCysLeuLeuTrpAlaThrPro--ThrProAlaTrpProSerGlyThrVal 623
QY 75 TCCAGACAGACACCATCCCGCC-----AGGAGACTACCGCTGCTGCGCATCCTACC 125
DB 624 ArgSerGlnAlaProHisProTrpLysTrpArgProSerSerCysLeuSerHisThrCys 643
QY 126 ACCACGGAGAGCTGCTCTTTCAGTTCACAGCTGAGCTGAG-----CTGGAGATGCT 179
DB 644 LysArgProAlaTrpSerProArgCysProHisArgMetLysGlnLeuLysTrpGlnPhe 663
QY 180 GTGAGAGCCATGCCAGCTGTCGGCCTTTGTGTGTCACCAACGAGCAGCAGTACGAGG 239
DB 664 GlnArgLeu-----SerLeuArgGlnArgLeuArgProArg***ArgCys 678
QY 240 AGCAGTGGGAGAGAGCCCTTCCAA-----GGGAGA 269
DB 679 GlySerSerArgLysProTrpArgArgCysSerProGlyArgAla***AlaGlyArg 698
QY 270 TGGCAGACCTCTCTGAGGTGATAGTAGATGATCCCGCATC-----GGAAGTCAGAG 323
DB 699 Trp-ArgProSerAlaArgThrThrArg--ThrSerProValAsnTyrAlaArgGln 717
QY 324 GGGGTGTAGGTGATGAGAGAGAGATACGTCTTCAAGCAGTCAATTAAGGAGA 383
DB 717 GlnGlyThrGlyThr***ArgHisThrSerGlySerCysArgSerGly----- 733
QY 384 ATGGTCTTGCTCCAGAGAGAGAGAGAGATCCAGCCCTG----- 420
DB 734 -TrpSerCysCys---ArgGlnArgGlnProGlnLeuSerArgGlySerProValProG 752
QY 421 -----TTACCTCTACCTCTGCCCCCAGAGTGGCAGCTGCTTTTTCAG 467
DB 752 yProArgGlnLeuLeuProLeu***MetAlaProArgProTrpLeuTrpAlaSerAla 772
QY 468 ACTGATGAGAGCCAGAGTGGTCTGATCCCAACAGACACATATGTAAGGCTCTGG 527
DB 772 g-Trp-TrpGlyGln----- 776
QY 528 TGACCTATCTGAGGGCTCGGCTGACCAAGTACTATCTCAGAGCTGGCTTGCTG 587
DB 776 ----- 776
QY 588 GAGGAGTACTGTCAGTGGCAGCAGTACGTACCTGGGAACCCCTGAGACAAAGCT 647
DB 777 -----AlaProCysThrAlaAlaThrCysCysSerLeu----- 787
QY 648 AACATCCACAGACAGACAGATGTGACAGACAA----- 680
DB 788 -----ProGlySerLeuGlyLeuAlaTyrArgArgArgPhe 799
  
```

```

QY 681 -----ACGTGCAATTAATCCAAATGTTAAATGATGATTACAGCCTAGCTATGGAGC 734
DB 800 ProCysSerCysSerProLeuPheCys-----LeuValPro-ProProTrpAla 816
QY 735 TGCTGGC-----TCTAGTCCAGAAATATGGGGGTATGACTGCTCTTC 779
DB 816 AlenGlyTrpTrpProThrProAlaAsnSerProGlnSerGlyAlaAlaGlnLysPro 836
QY 780 AACCTG---TGAGCTGTAAGCAGCTCAGCTAGTCTCCCACTGGAG----- 825
DB 836 AlaLeuProGlnPro***AsnCysLeuArgLeuArgGlyProValGlyArgLeuSer 856
QY 826 -----GCTGTGCCCCCTCT-----CG 842
DB 856 ArgGlyGlnHisArgSerArgAlaHisArgLeuProValHisAsnArgSerGlyArg 876
QY 843 GACGTTCCGTGGGACAGCCCATCATG----- 870
DB 876 ySerProProSerSerSerProValIleArgAlaArgProLeuAlaAlaGlyGlnGly 896
QY 871 -----TGTTCAATAGTGTGAGAAAG 890
DB 896 ValaGlySerAlaAlaAlaGlyLysArgLysSerLysGlyProCysSer----- 911
QY 891 TAGCTAAAGCCCTGCTGCTGCTGTCACATGCCACAGCGGCTGGGCTGGTGA 950
DB 912 -----ArgGlyCysCysCysCysCysCysCysCysCysCysCysCysCysCys 926
QY 951 GGAATATCATCTGTGAGTGTCTCTCAGCTTAAGCTTGACAGACAGACTTGGCGGGA 1010
DB 927 ---Asp-HisArgProPheLeuSerPheGlyGlnAlaAlaLeuThrTrpMetGly 945
QY 1011 TGCTCAGAGATGGGTGATTTCTGTAACCTGGGAGAGCTATCTCTGACCTCCGACGG 1070
DB 945 yLeuGlnAlaTrpGlnGlySerLysProGlyArgProCysSerLysLeu-----H 962
QY 1071 ACACCTCCAGCGCAGCCAGCCAGGGGTCA---GGGACAGAGGTGACACCTCAGCATGAGCCA 1127
DB 962 IAlaProProProIleValaGlySerGlnSerAlaLysLeuSerCysAla***ArgPro 981
QY 1128 AGACTGGGCTCAGGAGACAG-----TGCTGTGAGCCAGAGACTG----- 1169
DB 982 AlaLeuGlySerValTrpArgAspLeuCysLeuLeuThrArgGlnLysCysPheCysGln 1001
QY 1170 GGGCGGGGTGGGGCGGGGCGCTTCTGCTCATTTGCTTAAGTAAGAAAGCCTCAAGGA 1229
DB 1002 ThrArgPhePheGlyAspProAlaPro-ProSerSerLeuAlaLeuSerGlnProG 1021
QY 1230 GCCAAACCCAGGCTTCCCTCTCGAGTTGAATATCCAGAAATCTTTGACTTCTT 1289
DB 1021 nProAlaPro-----ProProSerAla----- 1028
QY 1290 GTTGGTTAAATTGTTATTTTGTAAATAATAATAATAATAGTTAATAATG 1344
DB 1029 -----ValTrpIlePheIleAspLeuValLeuArgLeuAlaAspArgLeu 1093
  
```

RESULT 6  
 ABG14734 standard; Protein; 1040 AA.  
 ID ABG14734;  
 AC ABG14734;  
 DT 18-FEB-2002 (first entry)  
 DE Novel human diagnostic protein #14725.  
 DE Human; chromosome mapping; gene mapping; gene therapy; forensic;  
 KW food supplement; medical imaging; diagnostic; genetic disorder.  
 XX Homo sapiens.  
 OS  
 PN MO200175067-A2.

PD 11-OCT-2001.  
 XX 30-MAR-2001; 2001W0-US08631.  
 PR 31-MAR-2000; 2000US-0540217.  
 PR 23-AUG-2000; 2000US-0649167.  
 XX (HYSE-) HYSEQ INC.  
 PA Dmanac RT, Liu C, Tang YT;  
 XX WPI; 2001-639362/73.  
 DR N-PSDB; AAS78921.  
 XX  
 PT New isolated polynucleotide and encoded polypeptides, useful in  
 PT diagnostics, forensics, gene mapping, identification of mutations  
 PT responsible for genetic disorders or other traits and to assess  
 PT biodiversity

PS Claim 20; SEQ ID No 45093; 103bp; English.

XX The invention relates to isolated polynucleotide (I) and  
 CC polypeptide (II) sequences. (I) is useful as hybridisation probes,  
 CC polymerase chain reaction (PCR) primers, oligomers, and for chromosome  
 CC and gene mapping, and in recombinant production of (II). The  
 CC polynucleotides are also used in diagnostics as expressed sequence tags  
 CC for identifying expressed genes. (I) is useful in gene therapy techniques  
 CC to restore normal activity of (II) or to treat disease states involving  
 CC (II). (II) is useful for generating antibodies against it, detecting or  
 CC quantitating a polypeptide in tissue, as molecular weight markers and as  
 CC a food supplement. (II) and its binding partners are useful in medical  
 CC imaging of sites expressing (II). (I) and (II) are useful for treating  
 CC disorders involving aberrant protein expression or biological activity.  
 CC The polypeptide and polynucleotide sequences have applications in  
 CC diagnostics, forensics, gene mapping, identification of mutations  
 CC responsible for genetic disorders or other traits to assess biodiversity  
 CC and to produce other types of data and products dependent on DNA and  
 CC amino acid sequences. ABG00010-ABG0377 represent novel human  
 CC diagnostic amino acid sequences of the invention.  
 CC Note: The sequence data for this patent did not appear in the printed  
 CC specification, but was obtained in electronic format directly from WPI  
 CC at ftp.wipo.int/pub/published\_pct\_sequences.

XX Sequence 1040 AA;

Alignment Scores:

Prod. No.: 0.000142 Length: 1040  
 Score: 144.50 Matches: 109  
 Percent Similarity: 30.77% Conservative: 27  
 Best Local Similarity: 24.66% Mismatches: 134  
 Query Match: 5.64% Indels: 172  
 DB: 22 Gaps: 25

US-09-989-919-15 (1-1397) x ABG14734 (1-1040)

QY 1176 CCGGCCCGAGGCTCTGAGCTCAACACACCTGCTCCCTGAGCTTGAGTACGCT 1117  
 DB 520 ProAlaThrProGlyCysGlyAla-----SerProGlyProLeuAlaSerAla 535  
 QY 1116 GAGGTGTGACCTCTGCCCCCTGACCCCTGGGCTGGGAGGTGTCCCTGTGGGAGG 1057  
 DB 536 -----GlyPro\*\*AlaLeu\*\*GlyThrProSerSer----- 546  
 QY 1056 TCAGATAGACCTCCCGAGTACAGATACACCATCTGGAGCATCCCGCCAGAGTC 997  
 DB 547 AlaAlaThrSerGlnProGlyThrArgProSerSerSerTrpAlaProProCysGlnAla 566  
 QY 996 TCCTGTCCAGACCTAAGACTGAGAGAACATCTCACGATGTGTCCCGACGAGGCCCA 937  
 DB 567 ProCysPro\*\*ArgProAlaAlaSerProGlyMetGlyAlaProSer----- 583  
 QY 936 CCGCTGTCTGTGAGCATGTGACGAGCAGCAGAGGGGCTTTAGCTACATTCTTCACATA 877

DB 584 -----SerCysProAlaAlaSerAlaGly----- 591  
 QY 876 TGAACACAGTATGAGGCTGCCCGACGAAACCTCCAGGAGGCGACACCCCGCATGG 817  
 DB 592 -----ProProHisSerThrAlaGlyAlaHisSerProThrGly 604  
 QY 816 GGAGACTAGCTGAGCTTCTTACAGCCACAG-----GGTTGAGA 775  
 DB 605 TrpProSerProAlaCysAlaAlaGlyProGlnCys\*\*Thr\*\*GlySerGlyTrp--- 623  
 QY 774 GGCAGTCATACCCCGCATGATTCCTGAGTACAGGACGAGCCATGCTAGCTGCTGA 715  
 DB 624 ---SerCysTrpPro-----GlyArgSerPheProAla----- 633  
 QY 714 AACTGACATTTTAACTTTGCAATTTGCAATTTGCTGCTGCTGCTGCTGCTG 655  
 DB 634 ---SerLeuPheGlySerTrpProAlaCysPro---AlaSerProArgSerSerAlaGln 651  
 QY 654 GGATGT----- 649  
 DB 652 GlyCysLeuProLeuAlaAlaArgMetIleValAlaTrpProAlaProAlaAlaGly 671  
 QY 648 -----TAGCTTTGCTGTCAG-----GGGTTCCCGAGGTGAC 619  
 DB 672 ProGlnPro\*\*ThrProTrpProCysLeuArgArgGlySer\*\*GlyLeuSerArgSer 691  
 QY 618 ATGCAGTCTGTCAGTGCAG-----TCACTC-----CCTCCACAG 583  
 DB 692 TyrGluAlaGlyThrCysArgArgHisSerLeuSerAlaGlyGlyGlyHisProProSer 711  
 QY 582 GCAGGCCGAGCTGAGGATAGTACAGCTGTCAGCCGAGCCCTCAGATAGTGCAGCCAG 523  
 DB 712 ProAlaProLeuAlaGly-----ValArgSer 720  
 QY 522 AGCCCTTCATATGATGTGCTTGTGGATCAGGACCACTTGCTCC----- 475  
 DB 721 ProProSerAlaThrTrp-----ProGlyProProGlyThrsSerSerLeuHis 736  
 QY 474 -----ATCAGTCTTGAAMAAGACCAGT-----GCCGACTGG 442  
 DB 737 LeuProIleArgSerLeuProArgThrAla\*\*GlyProAlaSerAlaProAlaAspArg 756  
 QY 441 GGGGGCAGAGGTGAGAGGTAACAGAGGCTGAGATCTTCTTCTGAGGACAGACCATTC 382  
 DB 757 AlaGlyCysSerGlnProGlyAspGlyHisSer---AlaProLeuThrAlaArgPro--- 774  
 QY 381 TCCCTAATTGACTGCTTGAAGACAGCATATACCTCTCTCTCATCAGCTCAGACCCCT 322  
 DB 775 -----AlaAlaProArg 778  
 QY 321 CTGACTTCGATGGGAGTACTATCTATCTCAACCTCCAGAGGCTCTGCTCCCTT 262  
 DB 779 MetProProGlnProGlySer-----SerLeuGlyAlaAlaProAlaLeuHisAla 795  
 QY 261 GGAAGGGCTTCTCCACAGTGGCTCACCCTGTCAGAGGTGCTGTGTGTGACCAAGAGCC 202  
 DB 796 SerProAlaGlyProAlaArgGlnPro-----AlaVal-ArgPr 808  
 QY 201 CGACACTGGGAGTGGCTCTCAACAGATCCACAGCTCAGCCAGCT----- 156  
 DB 808 o---ProGlySerProAlaHisArgArgProAlaProGlnProGlyPheProProSerAl 827  
 QY 155 -----TGAACTGMAAAGAGGACCTCCCGTGGTGTAG 121  
 DB 827 AProTrpArgProArgLeuSerProArgProValGlyHisGlySerAlaThrProGlySe 847  
 QY 120 GATGCCAGCAGAC-----GGAGGCTTCGGG 94  
 DB 847 rAlaAlaGlnThrTrpSerAlaGlyProArgGlnHisProGlyThrAlaProGln 867  
 QY 93 GGGA 90  
 DB 867 yGly 868

XX	AB11908	standard; peptide; 1435 AA.
XX	AB11908;	
XX	11-JAN-2002	(first entry)
XX	Human BCL9 homologue,	SEQ ID NO:2178.
XX	Human; cytokine; cell	proliferation; cell differentiation; growth factor;
XX	haematopoiesis regulation;	tissue growth; immunomodulator; activin;
XX	inhibin; chemotaxis; chemokinesis;	thrombolysis; oncogenesis;
XX	proliferation; metastasis; cancer;	tumour; haematopoietic disorder;
XX	myeloid cell disorder; lymphoid	cell disorder; asthma; arthritis;
XX	chronic inflammatory condition;	proliferative retinopathy;
XX	atherosclerosis; coronary heart	disease; arterial ischaemia;
XX	bone disorder; osteoporosis;	vascular growth disorder;
XX	tissue regeneration; wound healing;	infection; immune disorder;
XX	cell culture; drug screening;	gene therapy; antiinflammatory;
XX	antiautismic; antiarthritic;	haemostatic; antiarteriosclerotic;
XX	cytostatic; osteoprotic;	vasotropic; cardiant; viroinide; antibacterial;
XX	antifungal; vulnerary;	antulcer.
XX	Homo sapiens.	
XX	WO200157188-A2.	
XX	09-AUG-2001.	
XX	05-FEB-2001;	2001WO-US03800.
XX	03-FEB-2000;	2000US-0496914.
XX	27-APR-2000;	2000US-0560875.
XX	(HYSE-) HYSEQ INC.	
XX	Tang YF, Liu C, Drmanac RT;	
XX	WPI, 2001-457740/49.	
XX	N-PSDB; ABA09052.	
XX	Human proteins and DNA	encoding sequences useful for preventing,
XX	treating or ameliorating a	medical condition in a mammalian subject
XX	e.g. arthritis and cancer.	
XX	Claim 20; Page 256-257;	1963pp; English.
XX	Sequences AB10981-AB12330	represent 1350 novel human polypeptides, and
XX	sequences ABA08225-ABA09574	represent nucleic acids encoding them. The
XX	invention also relates to	vectors and recombinant host cells comprising a
XX	nucleotide of the invention,	methods of producing the novel polypeptides,
XX	antibodies against the	polypeptides, methods of detecting the nucleotides
XX	or polypeptides in a sample,	and methods of identifying compounds which
XX	bind to polypeptides of the	invention. Although novel, many of the
XX	polypeptides of the invention	have homology to known proteins, thereby
XX	giving an insight into their	probable biological activities, and hence
XX	potential therapeutic	applications. The polypeptides of the invention may
XX	have various activities,	including cytokine, cell proliferation or cell
XX	differentiation activities;	stem cell growth factor activity;
XX	haematopoiesis regulatory	activity; tissue growth activity;
XX	immunomodulatory activity;	activin- or inhibin-related activities;
XX	chemotactic or chemokinetic	activities; haemostatic, thrombotic or
XX	thrombolytic activities;	receptor or ligand activities; or may be
XX	involved in oncogenesis,	cancer cell proliferation or metastasis.
XX	Depending on their	biological activities, polypeptides and nucleotides of
XX	the invention are useful	for preventing, treating or ameliorating medical
XX	conditions, e.g., by	protein or gene therapy. Such conditions include
XX	cancers, haematopoietic	disorders (e.g., myeloid or lymphoid cell
XX	disorders), chronic	inflammatory conditions (e.g., asthma or arthritis),
XX	proliferative retinopathy,	atherosclerosis, coronary heart disease,
XX	arterial ischaemia,	bone disorders (e.g., osteoporosis), and abnormal

	CC	vascular growth. Polypeptides involved may be used to promote wound repair (or nucleic acids encoding them) may be used to promote wound healing (e.g., of burns, incisions and ulcers), while those with immunomodulatory activities may be used in the treatment of viral, bacterial and fungal infections in addition to immune disorders.
	CC	Polypeptides with growth factor activity may be used in cell cultures to promote cell growth. For example, such polypeptides may be used to manipulate stem cells in culture to give rise to neuroepithelial cells that can be used to augment or replace cells damaged by illness.
	CC	autoimmune disease or accidental damage. The polypeptides and nucleotides may also be used in the diagnosis of the above conditions, and in drug screening techniques. The present sequence represents a novel human polypeptide of the invention.
	XX	
SQ	Sequence	1435 AA;
	Alignment Scores:	
Pred. No.:	0.000471	Length: 1435
Percent Similarity:	139.50	Matches: 117
Best Local Similarity:	30.47%	Conservative: 39
Query Match:	22.85%	Mismatched: 157
	22	Indels: 199
	Gaps:	25
US-09-989-919-15 (1-1397) x ABB11808 (1-1435)		
OY	1221 GGCTTCATTGAAAGCAATATGCAGGACAAAAGCGCCGCCACCACCCC-----	1174
Dd	810 Gllglnlhhsprrggnglin-----glutylgmetcglyproAgyProPheLeuPromet	827
OY	1173 GCCCAGGTCTTGGCTCA-----	1156
Dd	828 SerclnglyprrogiySerAhnserGlyleuArgAsnleuArgIuProIlleglyProAsp	847
OY	1155 -----AACACACTTGCTGCTCCCTGACCCCATCTTTG-----GCTCAT	1120
Dd	848 GlmArgThrAhnsenArAglyleuSerHismetProleuProleuAsnProsetSerahsn	867
OY	1119 GCTGAGGTGTGCACCTGTGCCCCCTGACCCCTGGGCTGCGTGGAGATGTCCCTGTGCGG	1066
Dd	868 ProthrSerleuAhnThrAlaProProValGlmlnrglyleuGlyArglyProleuAsp	887
OY	1059 AGTTCAGATAGCTCTCCACAGGTACAGATCACCCACATCTTGAGAGATCCCCGCCAA	1000
Dd	888 lleserValalaglySer-GlnValHisSer-----ProGlylleasnProleuLy	904
OY	999 GTCTCCCTGTCCAGACCTAAGCTGAGAGAACAACCTCAC-----	963
Dd	904 sSerPro-----ThmethisGlnValInsSerPrometLeuGI	917
OY	962 -----GATGATATGCTCCACAGGAGGCCACCGCTGCTGTGACATGTGCA	916
Dd	917 ySerProsercllyAsnleuLysserProGlnInthProserGlnleu-Alaglymetleua	937
OY	915 GCACGACGACAGGAGGGCTTAGTACTACATTCACACATATTGAACACAGTGTGGGCTGC	856
Dd	937 laglyProAlaAlaAlaAlaserile-----	945
OY	855 CCACGMAACCTCCACAGGAGGGGACACAGCCCCAGTGGGAGACTAGACCTGCT	796
Dd	946 -----LysSerProProval-----LeuGlySerAlaAla	956
OY	795 TACAGCCACAGGCTTGAGAGAGAGATCATACCCCATGTATCC-----TGACATAAGA	742
Dd	956 laserProValHisleuLysserProserleuProAlaProserProGlyTyPrphrses	976
OY	741 GCCAGACGTCCTCATAGCTAGCGCTGTAAACTCACATTTTAACATTTGGCATATTGCAAG	682
Dd	976 erProGluProPro-----	980
OY	681 TTtGTCTGTGTCATATCTGTCTGTCTGTGAGATGTAGCTTTGTCTGCAGGGTTTCCAGCT	622
Dd	981 -----LeuGlnSerProGlylleProProca	989

```

QY 621 GACATGACGTGCTGCCAGTCAAGTCACTTCCTCC---ACAGGCAAGCCAGCTGCTGAG 565
    |||
Db 989 snhlslsAlaProleuthrMetAlaSerProAlaMetleuGlyanValGlu----- 1006
QY 564 GATAGTCAGCTGCTGACGAGCCCTCAGATAGTACGACGAGGAGCCCTTACATATGTG 505
    |||
Db 1007 -----serGlyGlyProProProProProthralaSerGlnProAlaSerValAsnIleProG 1025
QY 504 TCTTGTGGATCAGGAGCACCATTGGCTC-----CATCCAGTCTTGAAGAAAGACC- 455
    |||
Db 1025 lYserleuProSerSerThrProTyrThrMetProGluProthrlleuSerGlnAsp 1045
QY 454 -----AGCTGCCAGCTGGGG 439
Db 1045 roleuSerlleuMetSerArgMetSerlySpheAlaMetProSerSerAsnProGlyT 1065
QY 438 GGCAGAGGTGAGAGTA----- 422
Db 1065 yrlshlslsAlaIlelyThrValAlaSerSerAspAspSerProProAlaArgS 1085
QY 421 -----ACAGGCTCGATGTTCT----- 404
Db 1085 erProAsnleuProSerMetAsnAsnMetProGlyMetGlyIleAsnThrGlnAsnProA 1105
QY 403 ---CTTCTGAGGAGCAGACCATT-----CTCCCTAATTGACTGCTTGAAGACAGTA 352
    |||
Db 1105 rglleSerGlyProAsnProValValProMetProthrlleuSerProMetGlyMetThrG 1125
QY 351 TACCTGCTCTCATCACTCAGACCCCTCTGATCCGATGGGGATCAGTATCTATC 292
    |||
Db 1125 lnrProleuSerhlSerAsnGlnMetProSerProAsnAlaValGlyProAsnIle---- 1143
QY 291 AACCTCCAGAGAGAGCTCTGCATCTCCCTTGAAGGAGCTTCTCCAGTGCACCTGTC 232
    |||
Db 1144 ---ProPro-----HlsglyValProMetG 1151
QY 231 CAGGTGCTGTGTGTGAGACCAAGGCGGACACTGGGCGATGGCTTCACAGACATCC 172
    |||
Db 1151 lYProGlyleuMet---SerhlS---AsnProIleMetGlyhlslGlySerGlnGluProP 1169
QY 171 ACAGCCTCAGCCAGCTTGAACACTGAAGAGGAGCAGCTCCCGTGT----- 126
    |||
Db 1169 oMetValProGlnGly-----ArgMetGlyPheProGlnGlyPheProProValG 1186
QY 125 -----GGTAGATGCGCAGACGCGTAGTCT 100
Db 1186 nserProProGlnGlnValProPheProhlSAsnGlyProSerGlyGlyGlnGlySerP 1206
QY 99 TCCTGGGGGATGTGCTGTGGGATACACTGTGACT-----CGTACTGCTG 52
    |||
Db 1206 eProGlyGly-----MetGlyPheProGlyGlnGlyProleuGlyYrgrProSerAs 1223
QY 51 CTGGCCGTGAGTTCTGACATGCTGCC 24
    |||
Db 1223 nleuProGlnSerSerAlaAspAla 1232

```

FT misc\_difference 43  
 FT /note= "Encoded by in-frame stop codon"  
 FT misc\_difference 192  
 FT /note= "Encoded by in-frame stop codon"  
 FT misc\_difference 258  
 FT /note= "Encoded by in-frame stop codon"  
 FT misc\_difference 43  
 FT /note= "Encoded by in-frame stop codon"  
 FT misc\_difference 374  
 FT /note= "Encoded by in-frame stop codon"  
 FT misc\_difference 397  
 FT /note= "Encoded by in-frame stop codon"  
 FT misc\_difference 452  
 FT /note= "Encoded by in-frame stop codon"  
 FT misc\_difference 458  
 FT /note= "Encoded by in-frame stop codon"  
 FT misc\_difference 475  
 FT /note= "Encoded by in-frame stop codon"  
 FT misc\_difference 487  
 FT /note= "Encoded by in-frame stop codon"  
 FT /note= "Encoded by in-frame stop codon"  
 PN WO200233093-A2.  
 XX  
 XX 25-APR-2002.  
 PD  
 XX 17-OCT-2001; 2001WO-US32458.  
 PF  
 XX 18-OCT-2000; 2000US-241469P.  
 PR  
 XX (GEMV ) GENETICS INST INC.  
 PA  
 XX  
 PI Racie LA, Twine NC, Agostino MJ, Wolfman NM, Morris EA;  
 XX  
 XX WPI; 2002-452389/48.  
 DR  
 XX N-PSDB; ABK52580.  
 PT Novel purified aggrecanase polypeptide useful for developing inhibitors  
 PT and antibodies to the aggrecanase polypeptide, which are useful for  
 PT treating aggrecanase-associated condition such as osteoarthritis  
 PS Disclosure; Page 39-41; 41pp; English.  
 XX  
 XX This invention relates to the CDNA and protein sequences of a novel  
 CC human aggrecanase polypeptide. The protein of the invention may be  
 CC used to inhibit the proteolytic activity of aggrecanase, or to inhibit  
 CC the aggrecanase-mediated cleavage of aggrecan in cartilage. The  
 CC protein of the invention is useful for developing inhibitors of  
 CC aggrecanase protein. The CDNA sequence encoding the aggrecanase protein  
 CC of the invention is useful for designing probes for obtaining DNA  
 CC sequences encoding other aggrecanase molecules. The CDNA sequence is  
 CC also useful for detecting mRNA encoding aggrecanase in a given cell  
 CC population, and thus for detecting or diagnosing genetic disorders  
 CC involving the aggrecanase, or disorders involving cellular, organ or  
 CC tissue disorders in which aggrecanase is irregularly transcribed or  
 CC expressed. The DNA sequences may also be useful for preparing vectors  
 CC for gene therapy applications. An inhibitor of the protein is useful  
 CC in treating conditions characterised by degradation of articular  
 CC cartilage, by blocking the enzyme's proteolytic activity. An  
 CC aggrecanase protein inhibitor and a method for inhibition of its activity  
 CC are useful for treating various aggrecanase-associated conditions  
 CC including osteoarthritis and other inflammatory diseases. The  
 CC present sequence represents a human Hsa011374 protein of the  
 CC invention.  
 CC  
 CC  
 SQ Sequence 738 AA;  
 XX  
 XX

Alignment Scores:  
 Pred. No.: 0.00109  
 Score: 134.50  
 Percent Similarity: 32.09%  
 Best Local Similarity: 24.11%  
 Query Match: 5.25%  
 DB: 23  
 Length: 738  
 Matches: 136  
 Conservative: 45  
 Mismatches: 170  
 Indels: 214  
 Gaps: 29

US-09-989-919-15 (1-1397) x AAU7643 (1-738)

QY 1265 TTCAAACTCGAGAGAGGAGAAAGCTGTTGTTGCTTTGAGGCTTTTCATTGAAGC 1206  
 DB 2 PheGlyLeuGlnHisArgIleProGlySerGlyCysGlyProSerGlyHisValMet 21  
 QY 1205 AATATGAGGAGAAAGGCGCGCGCCAGCCCGCCAGGCTCT-----GGCT 1158  
 DB 22 AlaSerGlnArgIleArgIleProAlaPro--AlaSerProGlyProProAlaAlaGlyS 41  
 QY 1157 CAACACACACCTGCTCTGAGCCCACTGCTGCTATAGCTGAGGTGTC-----A 1107  
 DB 41 erCys\*\*AlaCysSer--AspProSerLeuArgIleArgIleSerLeuGlyTyrProPro 60  
 QY 1106 CCTGTGCGCTGACCCCTGCGCTGCG-----CTGG 1077  
 DB 60 HisSerAlaPro-----AlaGlyAlaLeuValLeuValProAlaLysSerArgLeu 77  
 QY 1076 GAGTGTCCCTGTGCGGAGGTGAGATAGCTCCCGAGGTACAGATCACCACATCTCT 1017  
 DB 77 euValGlyGlyAlaGlyArg--GlnLeuLeuPheProLeuThrIleGlyHisAlaSerI 96  
 QY 1016 GAGAGCATC--CCCCGCAAGTCTCTGTCCAGACTAGCTAGAGAAACATCCACGAT 960  
 DB 96 ysArgPheHisProArgAlaHisSer----- 104  
 QY 959 GAGTGTCCCGCCAGAGCCCGCCAGGCTGCTG----- 925  
 DB 105 -----SerValProProProProGlyValHisProGlyThrGluProGlyLeus 121  
 QY 924 ----GCATGTGAGCAGCAGCAGAGGGGCTTTAGCTACATTCTCACAATTATGAAC 870  
 DB 121 erArgIleLeuSerGlnArgMetThrGlyAlaLeuVal----- 133  
 QY 869 AGTGAATGGGCTCCCAAGG-----AACGTCCTCCAGGAGGGGAC 828  
 DB 134 ----TyrAspProProArgProGlnProGlySerAlaGlyHisProArgAsnAlaHisI 152  
 QY 827 GCGCCAGTGGGAGACTAGCTGAG-----CTTGCTACAGCCCAAGGCT----- 781  
 DB 152 euGlyLeuTyrTyrSerAlaAsnGlnGlnCysArgValAlaPheGlyProLysAlaVal 172  
 QY 780 -----TGGAGAGCAGCATATACCCCATGATTCCTGAGCTAGGA----- 742  
 DB 172 lacYthrPheAlaArgGlnHisLeuValSerLeuProAlaValAlaTyrAspTyrLeu\* 192  
 QY 741 -----GCCAGCAGTCCCATAGCTAGGCTGTAACATTCATTTAATTTGACATT 690  
 DB 192 \*\*GlyProSerAlaSerProSerSerArgProProLys-----ArgAlaTyr-Ile 208  
 QY 689 ATTGCACGTTTTCCTGTCACATCTGCTGCTGCGGATTTAGCTTGTCTGACAGGGT 630  
 DB 209 CysAlaArgProSerProAlaThrGlnThrArgTyr----- 220  
 QY 629 TCCGAGTGCATGCTGCTGCGAGTCAAGTCACTCCCTCCACAGGCAAGCCAGCTG 570  
 DB 221 -----ThrIysAlaAlaAlaAlaAlaSerSerPheLeuSerTyrPheGlyGlnAsn 237  
 QY 569 CTG-----AGGATAGTCAGC-----TGTTCAGCCGAGCCCTCA 537  
 DB 238 ValAlaTyrPargSerGlyAlaProArgValAlaAlaAlaProThrPsr\*\*ProPro 257  
 QY 536 GATAGTCAAGCAGAGCCCTTACATATGTGCTTGTGGATCAGGAGCACTTGGCT 477  
 DB 258 \*\*GlnGlnCysMetGly-----AlaGlyLeuAlaGlyValProGlyValLeuAla 274  
 QY 476 CCATCCAGTCTTAAAAAGACAGCTCCGAGCTGGGGGAGAGGTGAGAGGTAAACAG 417  
 DB 275 ProAlaProAlaGlnGlnValTyrPsrPro-GlyGlyGly-----SerAlaThr 291  
 QY 416 GCTGATGTTTCTTCTTCTGAGGCAAGACATTCTC----- 380

DB 291 rProAspLeuProLeuGlyGlyValHisValLeuValLeuThrSerArgProArgCys 311  
 QY 379 -----CTAATT 372  
 DB 311 aThrIleuArgProAlaArgArgProSerTyrPsrSerCysArgAsnSerAlaProGly 331  
 QY 371 GAGTGTGAGAGACAGATATCTCTCTCT----- 341  
 DB 331 oThrAlaSerArgCysAlaProProLeuAlaAlaProProSerThrThrGlyValLeu 351  
 QY 340 ----CATCACTCAGACACCCCTGACTCCGAT----- 310  
 DB 351 uThrHis--ThrAlaLysGlyMetLeuCysAlaAspThrCysAlaGlyProLeuAlaArg 371  
 QY 309 -----GGGGATCACTATCATCACTCCAGAGAGGCTCGCC-----ATCTCC 264  
 DB 371 lAsSerSer\*\*\*SerValGlnThrAlaSerSerMetGlyProGlyValCysGlnValAla 391  
 QY 263 TTGGAAGGCTTCTCTCC-----ACTGCTCAGCTGTCCAG 228  
 DB 391 roGlyArgThrGlyPro\*\*AlaCysValCysArgAlaAlaGly-HisLeuAlaVal 410  
 QY 227 TGTCTGCTGTGACCAAGAGCCCGACACTGG----- 192  
 DB 411 MetValGlyTyrThrProSerArgTyrGlyThrGlyAlaArgCysValValGlyThr 430  
 QY 191 -----CATGCTCT----- 183  
 DB 431 AlaArgAlaAlaHisGlyArgAlaLeuSerGlnLeuAlaGlnArgGlnAsnMetSerArg 450  
 QY 182 ---CACAGCATCCACAGCTCAGCCAGGTTGAACACTG-----AAAGAGGAGCTCC 132  
 DB 451 Phe\*\*GlnLeuProProThr\*\*ProValSerThrLeuProThrThrGlyLeuSerSer 470  
 QY 131 CGTGTGTGATGATGCGCAGCAGCGTACT----- 100  
 DB 471 HisThrTyrArg\*\*GlySerGlnGlyAlaMetSerTyrLeuGlyArg\*\*AlaSerPro 490  
 QY 99 -----TCCTGGGGAGTGTGCTGCTGCTGAGTATACACTGTACTG 61  
 DB 491 LeuThrProProThrProProSerThrPargMetValValSerSerThrGluThrProSer 510

RESULT 9  
 ID AAU32775  
 18-DEC-2001 (first entry)  
 AAU32775;  
 Novel human secreted protein #3266.  
 Human; vaccination; gene therapy; nutritional supplement;  
 stem cell proliferation; haematopoiesis; nerve tissue regeneration;  
 immune suppression; immune stimulation; anti-inflammatory; leukaemia.  
 Homo sapiens.  
 W0200179449-A2.  
 25-OCT-2001.  
 16-APR-2001, 2001MO-US08656.  
 18-APR-2000, 2000US-052929.  
 26-JAN-2001, 2001US-0770160.  
 (HYSE-) HYSEQ INC.  
 Tang YT, Liu C, Dmanac RT;  
 WPI, 2001-611725/70.



FT Region 622..630  
 FT /note="This line is omitted in the specification"  
 FT 883..891  
 FT /note="This line is omitted in the specification"  
 XX  
 PN JP06014783-A.  
 XX  
 PD 25-JAN-1994.  
 XX  
 PF 30-JUN-1992; 92JP-0173155.  
 XX  
 PR 30-JUN-1992; 92JP-0173155.  
 XX  
 PA (MITU) MITSUBISHI KASEI CORP.  
 XX  
 DR WPI: 1994-061478/08.  
 XX N-PSDB; AA056916.  
 XX  
 PT New glutamic acid receptor and gene - for use in analysis of  
 PT synaptic signal transduction, and diagnosis of brain disease  
 XX  
 PS Claim 1, Page 25-31; 35pp; Japanese.  
 CC  
 CC The sequence shows a glutamic acid receptor. The receptor is  
 CC useful for the analysis of nerve signal transduction; within the  
 CC synapse, expression of synapse plasticity, nerve cell necrosis,  
 CC brain structure and brain disease. It can also be used in gene  
 CC therapy.  
 CC  
 SQ Sequence 1239 AA;

Alignment Scores:  
 Pred. No.: 0.0128 Length: 1239  
 Score: 124.00 Matches: 94  
 Percent Similarity: 30.63% Conservative: 23  
 Best Local Similarity: 24.61% Mismatches: 112  
 Query Match: 4.84% Indels: 153  
 DB: Gaps: 24

US-09-989-919-15 (1-1397) x AAR45945 (1-1239)

QY 1196 AGAAGAGCCCGGCGCCCGCCCGCCAGCTCTGCTCAACACACACCTGCTCCCTGA 1137  
 DB 927 ArgArgAlaProAlaPro-----ThrThrSerGlyPro----- 937  
 QY 1136 CCCCACTTGGCTCATGCTGAGCTGCACTCTGCCCCCTGACCCCTGGG----- 1086  
 DB 938 -----ArgSerCysThrProGlyPro-----ProGlyGlnProSer 949  
 QY 1085 ---CTGGCTGGAGAGTGTCCCTGTGTGGGAGGTGAGATAGCTCCCGCAGTACAGAT 1029  
 DB 950 ProSerGlyTyrPargProPro-----GlyGlyGlyArg-----ThrProLeuAlaArgArg 966  
 QY 1028 CACCCACATCTGAGAGCATCCCGCCAGTCTCTGTCAGACCTAAGCTGAGAGACA 969  
 DB 967 AlaProGlnProProAlaArgPro-----GlyProAlaGln----- 978  
 QY 968 CTCACAGATGATTTGTCACAGAGCCCGCCAGCTCTGTGGCATGTGCAGACGAG 909  
 DB 979 -----GlyArgLeuSerProThrCysProGlnHis----- 988  
 QY 908 CAGCAGGGGCTTAAAGTACATCTCACTATTAACACAGTATGAGGGGCTGCCACGGA 849  
 DB 989 -----ProAlaGlyThrLeuGlyMetArgGlyGlnCysGluSer 1002  
 QY 848 ACCGTCACAGGAGGAGGAGCAGCCCGCAGTGGGAGAGTACCTG-----AGCTTG 798  
 DB 1003 GlyIleArgSerArgThrSerArgProProGluArgAlaLeuProGluArgSerLeu 1022  
 QY 797 CTTACAGCCCAAC-----AGGTTGAGAGGAGCAGTC 768  
 DB 1023 LeuHisIleHisCysHisItyrSerSerPheProArgIleGluArgSerGlyArgProPhe 1042

QY 767 ATACCCCATGATTCCTGAGTACAGACCCAGCA-----GTCCCATAGTACGCC 720  
 DB 1043 LeuPro-----LeuPheProGluProProGluProAspAspLeuProLeuGly 1059  
 QY 719 TGGTAACCTCACATTTTAACATTTGGCATTAATTCACGTTTGCTGCTGACATCTGCT 660  
 DB 1060 ---ProGluGlnLeuAlaArgArgGluAlaLeuAlaAlaAlaAlaAlaAlaAla 1075  
 QY 659 GTCTGGAGATGATGATTTGTCTGACAGGGGTCCAGG---TGACATGAGTGTCCAGT 603  
 DB 1076 -----ArgGly-ProArgProArgHisAlaSerLeuProSe 1087  
 QY 602 GCAGATC-----ACTCCCTCCACAGGACCCAGC---TGCTGAGATAGTC 558  
 DB 1087 rSerValAlaGluAlaPheThrArgSerAsnProLeuProAlaArgCysThrGlyHisI 1107  
 QY 557 AGCTGTGACCGGAGCCCTCAGATAGGTACGCCAGAGCCCTTCATATGTGCTTGT 498  
 DB 1107 ACysAlaCysProCysProGln----- 1114  
 QY 497 GGGATCAGGAGCACCTTGCTCCATCCAGTCTTGAAAAGACCAGCTGCCGA----- 446  
 DB 1115 -----SerArgProSerCysArgHisValAl 1123  
 QY 445 -----CTTGGGGGAGAGAGTGAAGTAAACAGAGGCTGATG 408  
 DB 1123 aglnThrGlnSerLeuArgLeuProSerIyrArgGluAlaCysValGluGlyValProI 1143  
 QY 407 TTCTCTTTGAGGAGCAGACCATCTCCCTAATTGACTGCTTGAAGACAGTATACC 348  
 DB 1143 aglyValAlaAla-----ThrTrpGlnPr 1151  
 QY 347 TCTCTCTCAT-----CACCTCAGACCCCTCTGACTTCGATGGGGGATCACTAT 297  
 DB 1151 oArgGlnHisValCysLeuHisThrHisThrHisLeuProPheCysTrpGlyThrValCy 1171  
 QY 296 CTATCACTCCAGAGAGGTCTGTCATCTCCCTGGAAGGGCTTCCCATGCTGCAC 237  
 DB 1171 sArgHisProProProCysSerSerHisSerProtrp----- 1183  
 QY 236 CTGTCCAGGTGTGTTGTTGTT-----GACCACAAAGGCCGACACTGGG 192  
 DB 1184 -----LeuIleGlyThrTrpGluProProSerHisArgGlyArgThrLeuGl 1199  
 QY 191 CARG 188  
 DB 1199 yLeu 1200  
 RESULT 11  
 ABG03717  
 ID ABG03717 standard; Protein; 505 AA.  
 XX  
 AC ABG03717;  
 XX  
 DT 13-FEB-2002 (first entry)  
 XX  
 DE Novel human diagnostic protein #3708.  
 KW Human; chromosome mapping; gene mapping; gene therapy; forensic;  
 KW food supplement; medical imaging; diagnostic; genetic disorder.  
 XX Homo sapiens.  
 OS  
 PN WO200175067-A2.  
 XX  
 PD 11-OCT-2001.  
 XX  
 PF 30-MAR-2001; 2001WO-US08631.  
 XX  
 PR 31-MAR-2000; 2000US-0540217.  
 PR 23-AUG-2000; 2000US-0649167.  
 XX  
 PA (HYSE-) HYSEQ INC.



XX Drmanac RT, Liu C, Tang YT;  
 XX MPI, 2001-639362/73.  
 DR N-PSDB; AAS67904.  
 XX  
 PT New isolated polynucleotide and encoded polypeptides, useful in  
 PT diagnostics, forensics, gene mapping, identification of mutations  
 PT responsible for genetic disorders or other traits and to assess  
 PT biodiversity -  
 XX  
 PS Claim 20; SEQ ID No 34076; 103bp; English.  
 XX  
 CC The invention relates to isolated polynucleotide (I) and  
 CC polypeptide (II) sequences. (II) is useful as hybridisation probes,  
 CC polymerase chain reaction (PCR) primers, oligomers, and for chromosome  
 CC and gene mapping, and in recombinant production of (II). The  
 CC polynucleotides are also used in diagnostics as expressed sequence tags  
 CC for identifying expressed genes. (I) is useful in gene therapy techniques  
 CC to restore normal activity of (II) or to treat disease states involving  
 CC (II). (II) is useful for generating antibodies against it, detecting or  
 CC quantitating a polypeptide in tissue, as molecular weight markers and as  
 CC a food supplement. (II) and its binding partners are useful in medical  
 CC imaging of sites expressing (II). (I) and (II) are useful for treating  
 CC disorders involving aberrant protein expression or biological activity.  
 CC The polypeptide and polynucleotide sequences have applications in  
 CC diagnostics, forensics, gene mapping, identification of mutations  
 CC responsible for genetic disorders or other traits to assess biodiversity  
 CC and to produce other types of data and products dependent on DNA and  
 CC amino acid sequences. ABG0010-ABG3037 represent novel human  
 CC diagnostic amino acid sequences of the invention.  
 CC Note: The sequence data for this patent did not appear in the printed  
 CC specification, but was obtained in electronic format directly from WIPO  
 CC at ftp.wipo.int/pub/published\_pat\_sequences.

XX Sequence 505 AA;

XX

Alignment Scores:  
 Pred. No.: 0.0102 Length: 505  
 Score: 123.50 Matches: 111  
 Percent Similarity: 33.56% Conservative: 36  
 Best Local Similarity: 25.34% Mismatches: 162  
 Query Match: 4.82% Indels: 129  
 DB: 22 Gaps: 23

US-09-989-919-15 (1-1397) x:ABG03717 (1-505)

QY 1196 AGAAGGCGCCG-----GCCCGACCCCGCCGAGTCTGCTCAACACGACCT 1146  
 DB 106 ArgGlyAlaProAlaThrGlyAlaProProProSerAlaAlaSerProThrAla 125  
 QY 1145 GCTCCCTGACCCGAG-----TCTTGCTCAAGCTGAGTGTGACCTCTG 1101  
 DB 126 ProProGlyProAlaThrGlyAlaArgValSerGlyAlaAlaAlaArg--AlaProPro 144  
 QY 1100 CCCCTGACCCCGGCTGCGCTGGAGCTGCCCTGCGGGAGGT----- 1056  
 DB 145 GlyThrAlaProAlaGlyGlyAlaArgArgProGlyGlyAlaProAlaThrGluPro 164  
 QY 1055 -----CAGAGATAGCTCCCGCAGGTACAGATACACCATCTGGAGAGA-----TTC 1008  
 DB 165 ProAspSerArgThrProAlaAlaAlaAlaAlaAlaSerSerAlaProGlyAlaValSerGly 184  
 QY 1007 CCCGCAAGTCTCTCTGTCAGACCTTAAGCTGAGAGAACTCC----- 965  
 DB 185 ProAlaAlaAlaPro-GlyProProGlyArgArgGluAlaAlaGlyArg\*\*ProG 204  
 QY 964 -----ACGATGATGTCGCCCGACCGACCGCCCG----- 938  
 DB 204 AspAlaGly\*\*ArgGlyLeuTrpGluGlyAlaLeuProValProGlySerSerProG 224  
 QY 937 -----ACGACCTGCTGTGTCATGTGCAGCAGACGAGGAGGCTTTGCTACATCT 885

DB 224 ThrSerSerSerSerThrGlyArgThrSerGlySerArgAlaProSerHisMet-- 243  
 QY 884 CACACTATTGACACAGAGTATGGGGCTGCCCGAGAACCGGCCGAGGAGGACAGCC 825  
 DB 244 -----ValProGlyThrGlySerProProGlyArgGlyGlyAlaG 258  
 QY 824 CCCAGTGGGAGACTAGCTGAGCTTGCTGACGCCACAGGAGTTGAGAGGACGATCAT 765  
 DB 256 ValArg\*\*AlaAlaAlaProAlaGlyValAlaSerSerSerLeu----- 273  
 QY 764 CCCCATGATTCCTGAGCTAGGAGCGACGATCCCATAGCTAGGCTGTAACTACATT 705  
 DB 274 -----TblyLys\*\*\* 277  
 QY 704 TTAACATTGGCATTTATGACGTTGT-----CCTGTCACATCTGTCT 660  
 DB 278 -----LeuAlaLeuPheArgProCysPheGlnGluProThrProGlySerValGlyC 295  
 QY 659 GTCCTGAGATGTTAGCTTGTCTGACG-----GGTTCCGAGGTGACATGC 615  
 DB 295 ArgGlyProLeuGlu-CysPheThrHisSerSerProValGlyValaSerGlyHisArgH 315  
 QY 614 AGTGC-----TGCAGTGCAGACTACTCTCCACAGGCAAGCCAGCTGAGAGATA 561  
 DB 315 IAspAspAsnCys---CysArgProLeuLysProProSerProLysAlaAla-TpAla 333  
 QY 560 GTCAGCTGTGACCGGAGCCCTCAGATAGTACGC-----AGA 522  
 DB 334 ValProArgAlaAlaValProLysAlaHisAla\*\*Lys\*\*ArgAlaAlaLysPheArg 353  
 QY 521 GGCCTTCACATATGTTGTTCTTGTGGATCAGGACACACTTGGCTTCATCCAGTTGAA 462  
 DB 354 GlyLeuArgValLeuGly-----ProAsnValThr 363  
 QY 461 AAAGACAGCTCCGACCTGCGGGGCGCAGAGTGAAGTAAAGGCGCTG----- 413  
 DB 364 LeuSerAsnProProThrArgGlyPheArg\*\*LeuGly-ThrGlyValProGlyPheG 383  
 QY 412 -GATGTTCTCTTTCGAGGCAAGACCATTCCTCCTAATTGACGCTTGAAGACAG 354  
 DB 383 AspProCysValAspSerGlyLeu\*\*ValGluGluGlyLeuCysProGluAlaSerAr 403  
 QY 353 TATACCTCTCTCATCATCAGCAGACCCCTCTGACTTCGATGGGAGATCATATCA 294  
 DB 403 GgIysenGlyGluArgAsnLysGlyThr-----TblyIlePro--Pr 417  
 QY 293 TCAACCTCGAAGAGTCTGCATCTCCCTTGAAGGCTTCTCCACT-----G 243  
 DB 417 ogInProProLeuArg-ProSerSerArgTrpLeuGlnLys\*\*ProThrProLeuProG 437  
 QY 242 GCTCAGCTCTCAGAGTGTCTGTGTGTGACACCAAGAGCCGACACCTGGCATGGCTCT 183  
 DB 437 LysPro\*\*Asp-----AlaThrSer-ProProAlaGlyGlyArg 451  
 QY 182 CACAGCATCCACAGCTCAGCAGGCTTGAACATGAAAGAGGAGGAGCTCCGTGTGT 123  
 DB 452 HisArgSerArgLeuProLysPro-----AlaLeuValGlyAsn----- 464  
 QY 122 AGATAGCGCAGAGCGGTAGTCTTCTGAGGAGATGGTGTCT 81  
 DB 465 -----AlaGlyThrSerSerLeuProAlaProGluProCys 476

RESULT 12  
 AAM87503  
 ID AAM87503 standard; Protein; 1212 AA.  
 XX  
 AC AAM87503;  
 XX  
 DT 23-FEB-1999 (first entry)  
 XX  
 DE Human N-methyl-D-aspartate receptor subunit encoded by clone NMDA22.  
 XX  
 KM Human; N-methyl-D-aspartate receptor; NMDAR2C;







XX  
DT 13-FEB-2002 (first entry)  
XX  
DE Novel human diagnostic protein #3524.  
XX  
OS Human; chromosome mapping; gene mapping; gene therapy; forensic;  
KW food supplement; medical imaging; diagnostic; genetic disorder.  
XX  
XX Homo sapiens.  
PN WO200175067-A2.  
PD 11-OCT-2001.  
PF 30-MAR-2001; 2001WO-US08631.  
PR 31-MAR-2000; 2000US-0540217.  
PR 23-AUG-2000; 2000US-0649167.  
XX  
PA (HSE-) HSEQ INC.  
PI Drmanac RT, Liu C, Tang YT;  
DR WPI, 2001-639362/73.  
XX N-PDSB: AAS67720.  
PT New isolated polynucleotide and encoded polypeptides, useful in  
PT diagnostics, forensics, gene mapping, identification of mutations  
PT responsible for genetic disorders or other traits and to assess  
PT biodiversity -  
PS Claim 20; SEQ ID No 33892; 103pp; English.

XX  
XX The invention relates to isolated polynucleotide (I) and  
CC polypeptide (II) sequences. (I) is useful as hybridisation probes,  
CC polymerase chain reaction (PCR) primers, oligomers, and for chromosome  
CC and gene mapping, and in recombinant production of (II). The  
CC polynucleotides are also used in diagnostics as expressed sequence tags  
CC for identifying expressed genes. (I) is useful in gene therapy techniques  
CC to restore normal activity of (II) or to treat disease states involving  
CC (II). (II) is useful for generating antibodies against it, detecting or  
CC quantitating a polypeptide in tissue, as molecular weight markers and as  
CC a food supplement. (II) and its binding partners are useful in medical  
CC imaging of sites expressing (II). (I) and (II) are useful for treating  
CC disorders involving aberrant protein expression or biological activity.  
CC The polypeptide and polynucleotide sequences have applications in  
CC diagnostics, forensics, gene mapping, identification of mutations  
CC responsible for genetic disorders or other traits to assess biodiversity  
CC and to produce other types of data and products dependent on DNA and  
CC amino acid sequences. ABG00010-ABG30377 represent novel human  
CC diagnostic amino acid sequences of the invention.  
CC Note: The sequence data for this patent did not appear in the printed  
CC specification, but was obtained in electronic format directly from WIPO  
CC at ftp.wipo.int/pub/published\_pct\_sequences.

XX  
SO Sequence 900 AA;

Alignment Scores:

Pred. No.:	0.0157	Length:	900
Score:	122.50	Matches:	96
Percent Similarity:	31.04%	Conservative:	44
Best Local Similarity:	21.23%	Mismatches:	156
Query Match:	4.79%	Indels:	156
DB:	22	Gaps:	19

US-09-989-919-15 (1-1397) x ABG03533 (1-900)

OY 1134 AAAGGCCCGGGCCCAACCCCGCCCAAGTCTGTGCCTAAACACACACTGCTCCTGACC 113BS  
Db :|||||  
489 ATGGTTPGTGGTCTGATGTCGACACTGTGCCCCCTGACCCCTGGGGTGCGCTGGGA 499  
OY 1134 CCAGTCTTGGCTCATGCTGAGGTGTCGACACTGTGCCCCCTGACCCCTGGGGTGCGCTGGGA 1075S  
:::|||||||

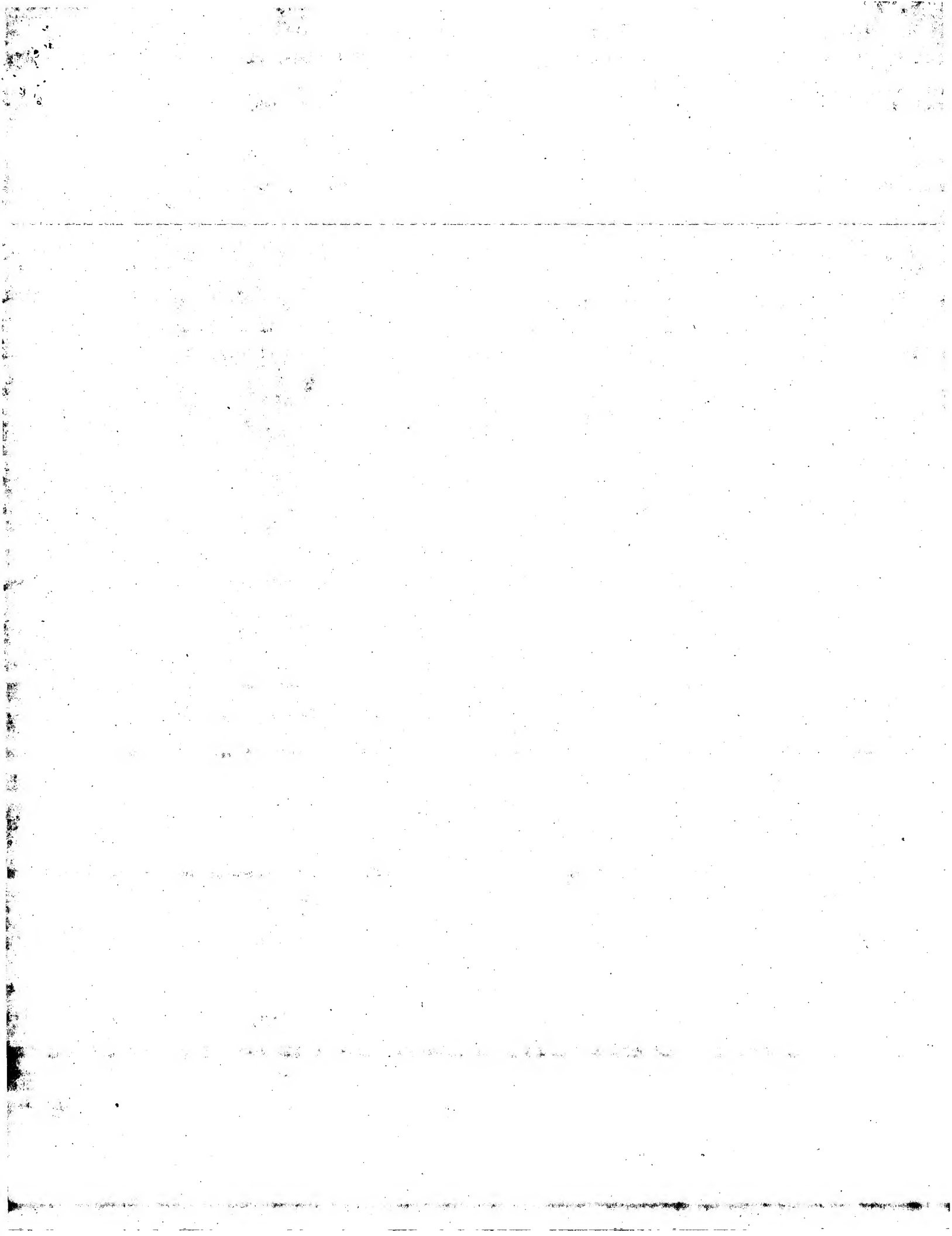
Db	500	-----	AlaIaIaProGlyArgGlyArgArgAlaGly	509
Oy	1074	GTGTCCCTGTCGGAGGTGACAGATACCTCCCGAGTACAGATACCCACATCTCTGG	101	
Db	510	SerAlaArgAlaArgGlyProGluProAlaProAlaProAlaArgAlaProGlyArgArgAla	529	
Oy	1014	AGCATCCCCCGGCAAGTCTCCGTCTCCAACTAAGCTGAGAACTCCACGA	961	
Db	530	ArgLeuProArgThr	541	
Oy	960	-----TGGATTGTCCCGACGAGCCCCACCGCTGCTGGCATGCGACGA	913	
Db	542	ProSerGluProTrpProLeuProSerArgProProGlySerGlyValArgGlyAlaSer	561	
Oy	912	GCAGCAGCAGGGGCTTAACTATTCACACTAATTGAACACAGTAGTGGGCTGCCA	853	
Db	562	AlaProProGly	568	
Oy	852	CGGAACCGTCCAGGAGGGGACAGCCCCCAGTGGGAGACTAGCTGAGCTTGCTTAC	793	
Db	569	-----SerArgSerGlyAsnSerProThrLeuGlnAspProGlyProHisSerGly	585	
Oy	792	AGCCACAGAGGTTTG	751	
Db	586	CysProGluGlyPheProArgLeuAlaGlyGluArgSerSerGlyProGluMetGly	605	
Oy	750	GCAGTAGAGCAGCAGCTCCATAGCTAGGCTGTAACTCACATTTAACTTTGGCAT	691	
Db	606	ThrValGlySerAsnAspValGlyIleAsnArgGlyValSer	624	
Oy	690	TATGTGACGTTGTCTGTGTACATCTGTCTGTGGATGTAACTTTGTCTGCAGGG	611	
Db	624	-----	624	
Oy	630	TTCCGAGTGATGACAGTCAAGTCCAGTCAAGTCACTCCCTCCAGCAGGCCAGCT	571	
Db	625	-----ValGlnArgThrTrpAlaGlyLeuLeuProProProGlySerMetGly	640	
Oy	570	GCTGAGATAGTCACTGTGTACGCGAGCCCTCAGATAGTCAACGACGAG	521	
Db	641	Ala--GlyThrGluTrpTyrGlnGlyThrProGluAsnProLysProArgGlyValArgAlaC	666	
Oy	520	-----GCCTTCACATATGTGGTCTTTGGATCAGGGACCACTTTGGCTCCATCAG	465	
Db	660	ValGlyCysSerPheThrTrpGlyAlaGlnAlaGlyIleValProGluTrpGlyLeuSer	680	
Oy	468	TCTGTAAAGACCAAGC	433	
Db	680	eralAlaLysIleLysLeuLysGlySerGlnAlaProSerAsnLeuCysGlnAlaGlyProIleAsn	700	
Oy	432	GGTGAAGGTAAACAGGGCTGATGTTTCTTTCTTGGAGGACAAACCATTTCCCTAAT	373	
Db	700	LeuLysIleGlnIleGlyIleArgSerMetProLysArgGlyGlyCysThr	716	
Oy	372	TGACTGCTTGAAGACAGTATACCTCTCTCATCAGCTGACAGACCCCTCTGACTTCC	313	
Db	716	YSerGlnCysHisValLysLeuGlnLeuGlnAlaLeuValGlyProGlySerThr	736	
Oy	312	GATGGGGGATCACTATTCATCAACTCAGAGAGGTCTGCACATCTCCCTTGAAGGGCT	253	
Db	736	LysTrp	741	
Oy	252	TCTCCACTGGCTCACTGTCCAGT	213	
Db	742	HisGlnSerSerValProGlyHisSerAlaProProGlyLeuIleLysGlySer	760	
Oy	212	-----CCAAAGGCCCGACACTGG	186	
Db	760	LysProIleAlaGlyProArgAlaProGlyCysGlyLeuArgLysGlnGlyGlyProG	780	
Oy	185	-----TTCACAGACATTCACAGCCTCAGCAG	158	
Db	780	LysLeu**AlaThrPheGlyValAlaProLeuSerGluAlaGlyValPro**GlySerPro	800	

QY 157 GTTGAACACTGAAGAGGACAGCTCCCTGTGTGTAGATGGCCAGACGGTACTCTTC 98  
 DB 800 rcaalallerprolysglyglnhisproglinprothrclyproserSerProglySerTPP 820  
 QY 97 CTGGG-----GGATGTCTGTCTGTGGATACCTGTACTCGTCTGTGT 50  
 DB 820 rogllyProarggluSerGlyTP-----ValSerValThrGly-----L 833  
 QY 49 TGCCGTGAGATTCTGCAGATACCTGCCCCG 21  
 DB 833 euProglinThrAlaAlaGlyProArgArg 842  
 RESULT 15  
 AAY44789  
 ID AAY44789 standard; Protein; 343 AA.  
 AC AAY44789;  
 AC AAY44789;  
 DT 04-MAY-2000 (first entry)  
 XX  
 DE "Corn Adenylsulphate kinase-2."  
 XX  
 KM Adenylsulphate kinase; Adenosine-5'-phosphosulphate kinase; APS kinase;  
 KM 3'-phospho-adenosine-5'-phosphosulphate; PAPS; sulphate assimilation;  
 KM corn; clone p0016.ctscj40rb; transgenic plant; screen; antibody.  
 XX  
 OS Zea mays.  
 XX  
 PN W0200004165-A1.  
 XX  
 PD 27-JAN-2000.  
 XX  
 PF 13-JUL-1999; 99MO-US15809.  
 XX  
 PR 14-JUL-1998; 98US-0092833.  
 XX  
 PA (DUPO) DU PONT DE NEMOURS & CO E. I.  
 XX  
 PI Falco SC, Allen SM, Anderson SL;  
 XX  
 DR WPI; 2000-182430/16.  
 DR N-PSDB; AAZ50160.  
 XX  
 PT New nucleic acid molecule and chimeric gene encoding an adenosine-5'  
 PT phosphosulphate kinase, useful for altering expression of sulfate  
 PT assimilation protein in plants  
 XX  
 PS Claim 6; Page 31-32; 42pp; English.  
 XX  
 CC The present sequence is the corn adenylsulphate kinase (APS kinase),  
 CC also known as adenosine-5' phosphosulphate kinase. This is obtained from  
 CC clone p0016.ctscj40rb, derived from corn pooled tassels shoots, p0016  
 CC cDNA library. This has 528 sequence identity to Catharantus roseus APS  
 CC kinase. APS kinase is a sulphate assimilation protein, that catalyses  
 CC the conversion of adenosine-5' phosphosulphate (APS) to 3'-phospho-  
 CC adenosine-5' phosphosulphate (PAPS). The nucleotide sequence is used as  
 CC probes and primers to identify, obtain and synthesize sulphate  
 CC assimilation proteins from other plants. It is also used to produce  
 CC transgenic plants, that are useful for altering the expression levels of  
 CC a sulphate assimilation protein. The APS kinase peptides are useful for  
 CC producing antibodies, that are used to screen and isolate cDNA clones.  
 XX  
 SQ Sequence 343 AA;  
 Alignment Scores:  
 Pred. No.: 0.0123 Length: 343  
 Score: 122.00 Matches: 88  
 Percent Similarity: 35.52% Conservative: 42  
 Best Local Similarity: 24.04% Mismatches: 118  
 Query Match: 4.77% Indels: 118  
 DB: 21 Gaps: 19

US-09-989-919-15 (1-1397) x AAY44789 (1-343)  
 QY 1182 CCCACCCCGCCCGACAGGCTCTGGCTCA-----AACACACTGCTCC 1141  
 DB 20 ProSerProAlaProGlyProAlaSerGlnGlyGlnArgGlnGlyAlaThrLeuLeuSer 39  
 QY 1140 CTGACCCCGAGCTGGGCTGAGGTGAGGTGACACTCTGCGCCCTGACCCCTGGGCTGGC 1081  
 DB 40 ProThrProThrLeuAla-ValIleLeuValAsnPro----- 51  
 QY 1080 CTGGAGAGTCCCTGTGGAGAGGTGAGAGTACAGATACCTCTCCAGATACAGATACACCA 1021  
 DB 52 -----GlnArgAlaProProValLeu----- 58  
 QY 1020 TCCTGGAGATCCCGCCCAAGTCTCTGTCTCAGACCTAACCTGAGAGACACTCCACGA 961  
 DB 59 -ProGlyLeuThrProSerSerAlaProLeuProAla-----LeuValIleHis-- 74  
 QY 960 TGGATTGTCGCCAGACGCCCGCCACCGCTCTGTGAGTGGAGTGGAGCAGC----- 908  
 DB 75 -GlyLeuThrProAlaGlySerSerHis-SerSerAlaGlyLeuAlaSerSerGlyArg 94  
 QY 907 -----AGCAGGGGCTTTAGCTACATTTCTCACTATTGAACACAGTATGGGAC 859  
 DB 94 rggluglygluglyargly-AlaArgThrHisCysHisArgGlyIleGlyArgTrp--- 112  
 QY 858 TGCCACAGGAAACCGTCCAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 799  
 DB 113 --ValArgArgArgArgArgArgArgArgArgArgArgArgArgArgArgArgArgArg 126  
 QY 798 GCTTACAGCCCGACAGGAGTGGAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 739  
 DB 127 ProHisSerProVal-----LysGluLysProValMetSerAsnIleGlyLys 142  
 QY 738 AGCAGTCCCAATGCTAGGCTGTGTAACCTCACTTTTAACTTTGCACTATTATGCACTTT 679  
 DB 143 SerThrAsnIle-----LeuThrIleAsnLysLeuIle 153  
 QY 678 GTTCGTGTACATCTGTCTGTCTG-----GGATGTACTTGTGTGAGGAGG 631  
 DB 154 GlyGlnSerAspArgGlnLysLeuLeuGlnGlnGlyGlyCysValValIleThrIle 173  
 QY 630 TTCACAGGTGACATGACAG-----TGTCGACAGTGCAGTCACTCCCTCCACAGGCA 580  
 DB 174 LeuSerGlySerGlyLysSerThrLeuAlaCys-AlaLeuSerArgGluLeuHisCysArg 193  
 QY 579 AGCCCACTGCTG-----AGCATGT 559  
 DB 193 ggllyHisLeuThrIleValLeuAspGlyAspAsnLeuArgHisGlyLeuAsnArgAsp 213  
 QY 558 CAGCTGTGACGCGAGGAGGCTCAGATAGTCA-----GC 526  
 DB 213 uSerPheLysAlaGlu-----AspArgAlaGlnAsnIleArgArgValGlyValAla 231  
 QY 525 CAGAGGCTTCACATATGTGTCTGTGGATGAGGAGCAGTGGCTCCATCCAGTCT 466  
 DB 231 AllysLeuPheAlaAspAlaGlyValIleCysIleAlaSerLeuIleSerProTyrArgArg 251  
 QY 465 TGAATAAGC-----CA 454  
 DB 251 gAspArgAspAlaCysArgAlaLeuLeuProHisSerAsnPheIleGluValPheIleAs 271  
 QY 453 GCTGCGACCTGGGGGAGAGGTGAGAGGTGAGAGGCTGATGTTCTCTTCTGAG 394  
 DB 271 pLeuProLeuLysIleCysGluAlaArgAspProLys---GlyLeuTyrIleLeuAlaArg 290  
 QY 393 GCAGAGCACTTCTCCTAATTGACTGCTTGAAGCAGCTATACCTCTCTCATCACC 334  
 DB 290 gThrGlyLysIleLysGlyPheThrGlyIleAspAspProTyr----- 304  
 QY 333 TCAGACACCCCTCTGACTCCAGTGGGAGTACACTATCATATCACTCCAGAGAGTCT 274  
 DB 305 -GluProPheIle-----AsnIleGluIleValIleLysMetLysAspGluIleLys 321

QY 273 GCCATCTCCC 264  
|||  
Db 321 sProSerPro 324

Search completed: June 24, 2003, 13:52:29  
Job time : 103.5 secs



GenCore version 5.1.6  
Copyright (c) 1993 - 2003 Compugen Ltd.

OM nucleic - protein search, using frame\_plus\_n2p model

Run on: June 24, 2003, 13:49:32 ; Search time 23 Seconds  
(without alignments)  
3574.245 Million cell updates/sec

Title: US-09-989-919-15

Perfect score: 2527

Sequence: 1 99tgcgtccactccgga.....aaaaaaaaaaagcggtc 1397

Scoring table: BLOSUM62  
Xgapop 10.0 , Xgapext 0.5  
Ygapop 10.0 , Ygapext 0.5  
Fgapop 6.0 , Fgapext 7.0  
Delop 6.0 , Delext 7.0

Searched: 262574 seqs, 29422922 residues

Total number of hits satisfying chosen parameters: 525148

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Command line parameters:  
-MODEL=frame\_plus\_n2p model -DBV=XLD  
-O=/cg2\_1/USFTO\_spool/US09989919/runat\_24062003\_102048\_19220/app\_query.fasta\_1.1543  
-DB=Issued\_Patents\_AA -QFMT=fastan -SUFFIX=rai -MINMATCH=0.1 -DOOPCL=0  
-DOOPEXT=0 -ONITS=bits -START=1 -END=1 -MATRIX=bloms62 -TRANS=human40.cdi  
-LIST=45 -DOCALLIGN=200 -THR SCORE=pct -THR MAX=100 -THR MIN=0 -ALIGN=15  
-MODE=LOCAL -OUTFMT=ptc -NORM=ext -HEAPSIZE=500 -MINLEN=0 -MAXLEN=2000000000  
-USER=US09989919 @CGN 1.1 28 @runat\_24062003\_102048\_19220 -NCPU=6 -ICPU=3  
-NO MAP -LARGEDUTERY -NEG\_SCORES=0 -WAIT -DSBLOCK=100 -LONELG  
-DEV\_TIMEOUT=120 -WARN\_TIMEOUT=30 -THREDS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOP=6  
-FGAPEXT=7 -YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database :

- 1: /cg2\_6/prodata/1/iaa/5A\_COMB.pep.\*
- 2: /cg2\_6/prodata/1/iaa/5B\_COMB.pep.\*
- 3: /cg2\_6/prodata/1/iaa/5A\_COMB.pep.\*
- 4: /cg2\_6/prodata/1/iaa/6B\_COMB.pep.\*
- 5: /cg2\_6/prodata/1/iaa/PCTUS\_COMB.pep.\*
- 6: /cg2\_6/prodata/1/iaa/backfile1.pep.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
C 1	124	4.8	1239	1	US-08-026-138E-3 Sequence 3, App1
C 2	120	4.7	1964	4	US-09-467-997-1 Sequence 1, App1
C 3	119	4.6	1063	1	US-08-093-453B-3 Sequence 3, App1
C 4	118	4.6	897	1	US-07-960-389-2 Sequence 2, App1
C 5	117	4.6	1185	4	US-09-041-886-23 Sequence 23, App1
C 6	116	4.5	433	4	US-09-046-158A-2 Sequence 2, App1
C 7	115	4.5	907	4	US-08-783-774-2 Sequence 2, App1
C 8	115	4.5	907	5	US-09-328-599A-1 Sequence 1, App1
C 9	115	4.5	907	5	PCT-US95-04611A-19 Sequence 19, App1
C 10	115	4.5	1063	1	US-08-127-499A-8 Sequence 8, App1
C 11	115	4.5	1063	1	US-08-482-847-8 Sequence 20, App1
C 12	112.5	4.4	376	4	US-08-874-569B-20 Sequence 20, App1

C 13	112	4.4	992	1	US-08-127-499A-1 Sequence 1, App1
C 14	112	4.4	992	1	US-08-482-847-1 Sequence 1, App1
C 15	109.5	4.3	1298	2	US-08-690-473-2 Sequence 2, App1
C 16	109.5	4.3	1298	4	US-09-259-821A-2 Sequence 2, App1
C 17	109.5	4.3	1298	4	US-08-843-655-2 Sequence 2, App1
C 18	109	4.3	1130	4	US-09-442-100-4 Sequence 4, App1
C 19	106	4.1	878	4	US-09-556-706B-2 Sequence 2, App1
C 20	105	4.1	557	4	US-08-927-218-139 Sequence 139, App
C 21	105	4.1	962	4	US-09-442-100-6 Sequence 8, App1
C 22	100.5	3.9	708	1	US-08-396-479B-8 Sequence 8, App1
C 23	100.5	3.9	708	1	US-08-818-823-8 Sequence 8, App1
C 24	100.5	3.9	739	1	US-08-396-479B-10 Sequence 10, App1
C 25	100.5	3.9	739	1	US-08-818-823-10 Sequence 10, App1
C 26	100.5	3.9	1068	1	US-08-396-479B-12 Sequence 12, App1
C 27	100.5	3.9	1068	1	US-08-818-823-12 Sequence 12, App1
C 28	100.5	3.9	1075	5	PCT-US94-07297-41 Sequence 41, App1
C 29	100.5	3.9	1214	2	US-08-231-193A-54 Sequence 54, App1
C 30	100.5	3.9	1214	2	US-08-486-273A-54 Sequence 54, App1
C 31	100.5	3.9	1214	3	US-08-480-474-54 Sequence 54, App1
C 32	100.5	3.9	1214	3	US-08-940-086A-54 Sequence 54, App1
C 33	100.5	3.9	1214	4	US-08-940-035A-54 Sequence 54, App1
C 34	100.5	3.9	1214	4	US-08-935-105A-54 Sequence 54, App1
C 35	100.5	3.9	1214	4	US-09-648-797-54 Sequence 54, App1
C 36	100.5	3.9	1219	2	US-08-231-193A-50 Sequence 50, App1
C 37	100.5	3.9	1219	2	US-08-486-273A-50 Sequence 50, App1
C 38	100.5	3.9	1219	3	US-08-480-474-50 Sequence 50, App1
C 39	100.5	3.9	1219	3	US-08-940-086A-50 Sequence 50, App1
C 40	100.5	3.9	1219	4	US-08-940-035A-50 Sequence 50, App1
C 41	100.5	3.9	1219	4	US-08-935-105A-50 Sequence 50, App1
C 42	100.5	3.9	1219	4	US-09-648-797-50 Sequence 50, App1
C 43	100.5	3.9	1231	2	US-08-231-193A-48 Sequence 48, App1
C 44	100.5	3.9	1231	2	US-08-486-273A-48 Sequence 48, App1
C 45	100.5	3.9	1231	3	US-08-480-474-48 Sequence 48, App1

#### ALIGNMENTS

RESULT 1  
US-08-026-138E-3  
Sequence 3, Application US/08026138E  
Patent No. 5502166  
GENERAL INFORMATION:  
APPLICANT: Masayoshi MISHINA  
TITLE OF INVENTION: NOVEL PROTEINS AND GENES CODING THE SAME  
NUMBER OF SEQUENCES: 19  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Nishionata Residence 1-107  
STREET: 5214, Nishionata-machi  
CITY: Niigata-shi  
STATE: Niigata-ken  
COUNTRY: JAPAN  
ZIP: 951  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Diskette, 3.50 inch, 1.44 MB storage  
COMPUTER: IBM Compatible  
OPERATING SYSTEM: MS-DOS V.5  
SOFTWARE: Word Perfect 5.1  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/026,138E  
FILING DATE: 26-FEB-1993  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: JP 39563/1992  
FILING DATE: 26-FEB-1992  
APPLICATION NUMBER: JP 173155/1992  
FILING DATE: 30-JUN-1992  
APPLICATION NUMBER: JP 215017/1992  
FILING DATE: 12-AUG-1992  
APPLICATION NUMBER: JP 303878/1992  
FILING DATE: 13-NOV-1992  
ATTORNEY/AGENT INFORMATION:  
NAME: Hamburg, C. Bruce  
REGISTRATION NUMBER: 22,389  
REFERENCE/DOCKET NUMBER: F-4551







LENGTH: 1063 amino acids  
 TYPE: amino acid  
 STRANDEDNESS: single  
 TOPOLOGY: linear  
 MOLECULE TYPE: peptide  
 HYPOTHETICAL: NO  
 ANTI-SENSE: NO  
 FRAGMENT TYPE: C-terminal  
 ORIGINAL SOURCE:  
 ORGANISM: Rubella virus  
 STRAIN: Therien  
 US-08-093-453B-3

Alignment Scores:  
 Pred. No.: 0.00158 Length: 1063  
 Score: 119.00 Matches: 96  
 Percent Similarity: 31.65% Conservative: 36  
 Best Local Similarity: 23.02% Mismatches: 136  
 Query Match: 4.65% Indels: 149  
 DB: 1 Gaps: 20

US-09-989-919-15 (1-1397) x US-08-093-453B-3 (1-1063)

QY 1184 GCCCAACCCCGCCCAAGCTCTGGCTCAACC-----ACACCTGCTCCCTGACCC 1134  
 DB 77 AlapProProProProGluArgGlnGlnThrArgSerGlnThrProAlaProLysPro 96  
 QY 1133 CAGCTTGCTCATGCTGAGGTGTCACCTGCTCCCTGACCCCTGGGCTGCTGGAG 1074  
 DB 97 SetArg-----AlaPro--ProGlnGlnProGln----- 105  
 QY 1073 TGTCCTCTGCGAGGTCAAGATACCTCCCGAGTACAG-----AAT 1029  
 DB 106 -----ProProArgMetGlnThrGlyArgGlyGlySer 116  
 QY 1028 CACCCACATCTGGAGATCCCGCCCAAGTCTCTGCTCCAGCTTAAGTGAAGA-- 972  
 DB 117 AlapProArgProGluLeuGlyProProThrAsnProGlnAlaAlaValAlaArgGly 136  
 QY 971 -----ACACTCCACAGATGATTGTCCCAAGCCCGCCCGCTGCTGT----- 926  
 DB 137 LeuArgProProLeuHisAspPro--AspThrGln-AlaProThrGlnAlaCysValTh 155  
 QY 925 ----GGATGTGACGACGACGACGAGGGCTTAAGTATCTCACTATTGAACAC 870  
 DB 155 rSerTrpLeuTrpSerGlnGlyGlnGlyAlaValPheTyrArgValAspLeuHisPheTh 175  
 QY 869 AGTATGGGGGTGCTCCCAAGGAACGCTCCCAAGGA----- 836  
 DB 175 rAsnLeuGlyThrProProLeuAspGluAspGlyArgTrpAspProAlaLeuMetTyrAs 195  
 QY 835 -----GGGCAAGCCCGCCCAAGTGGGAGACTAGCTGAGCTTGCTTAAGCCCAAG 783  
 DB 195 nProCysGlyProGluProPro--AlaHisValAlaArgAlaTyrAsnGlnProAlaG 214  
 QY 782 GTTGAAGAGCAGATCAACCCCATGATCTCTGACATGAGCAGACGAGATGCCATGCTA 723  
 DB 214 yAspValArgGlyValTrpGlyLysGlyLysArgThrTyrAlaGlnHisAspPheArgVa 234  
 QY 722 GGCTGTGAATCACTATTAACTATTGACATTTGACAGCTTGCTGCTGACATCTG 663  
 DB 234 lGlyGly-----ThrArgTrpHisArgLeu-----ArgMetP 246  
 QY 662 TCTGTCTGGATGTAGCTTTGTCTGACAGGGTTCCAGGTGACATGACGTGCTGCACT 603  
 DB 246 oValArgGlyLeuAspGlyAspSerAlaProLeuProPro--HisThrArgGluArg-- 265  
 QY 602 GCAAGTCACTCCCTCCACAGGACCCAGCTGCTGAGATAGTACGTGCTGACGG-- 545  
 DB 265 -eGluThrArg-----SerAlaArgHisProTyr 274  
 QY 544 -AGCCCTGAGATAGTACGACAGAGGCTTCACATATGTGCTTGTGGATCAGTACGAGAC 486  
 DB 544 -----

DB 274 pArgGLeuArgPheGlyAlaProGlnAlaPheLeuAlaGlyLeuLeuLeuAlaThrValAl 294  
 QY 485 CACTTGGCTCCATCCAGTCTTTGAAAAAGACCAAGCTCCGAGCTGGGGGCGAGAGTGAGA 426  
 DB 294 a-ValGlyThrAlaArgAlaGlyLeuGlnProArgAlaAspMet----- 308  
 QY 425 GGTAAACAGGGCTGATGTTTCTCTTTCTGAGGCAAGACATTCCTTAATTGACTGC 366  
 DB 308 ----- 308  
 QY 365 CTGAAAGACAGTAACTCTCTCTCATCACTACCTCAGACCCCTCTGACTTCCGATGGAG 306  
 DB 309 -----AlaAlaProProThr-LeuProGlnPro 317  
 QY 305 GATCACTATCTATCAACTCCAGAGAGGTCTGCATCTCCCTTGAAGGGCTTCTCCA 246  
 DB 318 -----Pro 318  
 QY 245 CTGGCTCACCTGTCCAGGTGCTGTGTGATACCAAGAGCCGACACTGGGATGAGC 186  
 DB 319 CysAlaHisGlyGlnHisTyrGlyHisHisHisGlnLeuProPheLeuGlyHisAsp 338  
 QY 185 TCTCAAGACATCCACAGCTTCAGCCAGGTGAACATGAAGAGGACGCTCCGTGCT 126  
 DB 339 GlyHisHis-----GlyGlyThrLeuArgVal 347  
 QY 125 GGTAGATGGCCAGCAGCGGTAGT-----CTTCCGTGGGGGATGG 87  
 DB 348 GlyGlnHisTyrArgAsnAlaSerAspValLeuProGlnHisTyrLeuGlnGlyTyr 367  
 QY 86 TGCTGTCTGGGA-----TACACTGTGTACTGGTACTGC 54  
 DB 368 GlyCysTyrAsnLeuSerAspTrpHisGlnGlyThrHisValCys 382

## RESULT 4

US-07-960-389-2  
 ; Sequence 2, Application US/07960389  
 ; Patent No. 5705611  
 ; GENERAL INFORMATION:  
 ; APPLICANT: HAYASHIDA, Kasuhiro;  
 ; TITLE OF INVENTION: Human GM-CSF Receptor Component  
 ; NUMBER OF SEQUENCES: 2  
 ; CORRESPONDENCE ADDRESS:  
 ; ADDRESSEE: Schering-Plough Corporation  
 ; STREET: 2000 Galloping Hill Road  
 ; CITY: Kenilworth  
 ; STATE: New Jersey  
 ; COUNTRY: USA  
 ; ZIP: 07033  
 ; COMPUTER READABLE FORM:  
 ; MEDIUM TYPE: Floppy Disc  
 ; OPERATING SYSTEM: System Software 7.1  
 ; SOFTWARE: Microsoft Word 5.1a  
 ; CURRENT APPLICATION DATA:  
 ; APPLICATION NUMBER: US/07/960,389  
 ; FILING DATE: 07-JAN-1993  
 ; CLASSIFICATION: 800  
 ; PRIOR APPLICATION DATA:  
 ; APPLICATION NUMBER: 554,745  
 ; FILING DATE: 18-JUL-1990  
 ; APPLICATION NUMBER: PCT/US 91/04846  
 ; FILING DATE: 16-JUL-1991  
 ; ATTORNEY/AGENT INFORMATION:  
 ; NAME: Blasdale, John H. C.  
 ; REGISTRATION NUMBER: 31,895  
 ; REFERENCE/DOCKET NUMBER: DX01430  
 ; TELECOMMUNICATION INFORMATION:  
 ; TELEPHONE: (908) 298-2902  
 ; TELEFAX: (908) 298-5388  
 ; INFORMATION FOR SEQ ID NO: 2:  
 ; SEQUENCE CHARACTERISTICS:  
 ; LENGTH: 897 amino acids



TYPE: amino acid  
 TOPOLOGY: linear  
 MOLECULE TYPE: protein  
 US-09-041-886-23

## Alignment Scores:

Pred. No.: 0.00265 Length: 1185  
 Score: 117.00 Matches: 92  
 Percent Similarity: 32.23% Conservative: 34  
 Best Local Similarity: 23.53% Mismatches: 140  
 Query Match: 4.57% Indels: 125  
 DB: 4 Gaps: 20

US-09-989-919-15 (1-1397) x US-09-041-886-23 (1-1185)

```

QY 1190 GCCCGGCGCCGACC----- 1176
DB 366 AlaProAlaProPheMetArgPheProTyrSerSerSerSerSerSerAlaAlaAla 385
QY 1175 -----CCGCCAGGTCTGCTCAACCA 1149
DB 386 SerSerSerSerSerSerSerSerSerSerAlaSerProPheProAlaSerGlnAlaLeu 405
QY 1148 CCGCT-----CCCTGACCCAGCTTGCTGCTCAGAGTGACCT 1104
DB 406 ProSerTyrProHisSerPheProProProProThSerLeuSerValSerAlaGlnProPro 425
QY 1103 CTGCGCCCTGACCCCTGGGCTGGCTGGAGTGTCCCTGCGGAGGTCAAGATGACCT 1044
DB 426 LysThrThrGlnProSerLeuProSerGlnAlaValTrpSerGlnGlyProProProPro 445
QY 1043 CCCCAG-----GTACAAATCACCCACATCTCTGAGATCCCGCCCAAG 999
DB 446 ProProTyrGlyAlaGluLeuAlaAsnSerAlaHisProGlyProPheProPro-- 464
QY 998 TCTCTCTCCAGACTTAAGCTAGAGAACTCCACGATGATGTCCCAAGCAGCC 939
DB 465 -----SerThrGlyAlaGlnSerThrAlaHisProProValSerThrHisHis 481
QY 938 CACCGCTGTGTGGCATGTGCACGACGACGAGGAGGCTTAACTACAT----- 888
DB 482 HisHisGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGln 501
QY 887 -----TTCACACTTAAACACAGTGAAGGCTGCC 855
DB 502 SerGlyProProProGlyAlaPheProHisProLeuGlu-----GlyGlySerSer 519
QY 854 CAGGGAACCGTCCAGAGGAGGAGGACAGCCCCAGTGGGAGACTAGCTGAGCTTCT 795
DB 520 HisHisAlaHisPro--TyrAlaMetSerPro-----SerLeuGlySerLeu 534
QY 794 ACACCCCAAGAGGTGAGAGGAGCACTATACCCCATGATCTTGAACCTAGAGCCAGCA 735
DB 535 ArgProTyrProProGlyProAlaHisLeuProPro----- 546
QY 734 GTCCCATAGCTAGGCTGTAACTCACTTTTAACATTGGCATTATTGACGTTTGC 675
DB 547 -----ProHis-----SerGlnValSer 552
QY 674 TGGTCACATCTGTCTGTCTGGATGTAGCTTTCTGTCAGAGGTTCCAGATGACATGC 615
DB 553 TyrSerGlnAlaGlyPro-----AaGlnGlyProProValSerSer 565
QY 614 AGTGTGCGCAGTGAAGTCACTCCCTCCAGAGCAAGCCAGCTGCTGAGATAGTACAG 555
DB 566 SerSerAlaSerSerSerSerSerThnSerGlnGlySerTyrProCys----- 580
QY 554 TGGTCAGCCGACCTCAG-----ATAGTCAAGCAAGGCTTACATATGTGT 504
DB 581 -SerHisProSerProSerGlnGlyProGlnGlyAlaProTyrProProProProValPr 600
QY 503 CTGTGTGATCAGGAGGACCACTGGCTCCATCAGCTTGAAGAAAGACAGCTGCCAGCC 444

```

```

DB 600 cThrValThrThrSerSerAlaThrLeuSerThrValIleAlaThrValAlaSerSerPr 620
QY 443 TGGGGGGAGAGGTCAGAGGTAACAGGGCTGATGTTCTTCTTCTGA----- 395
DB 620 AlaGlyTyrLys-----ThrAlaSerProProGlyProProProTyr 634
QY 394 -GGCAAGACATTCCTCAATTGACTGCTTGAAGACAGTAACTCTCTCATCA 336
DB 634 rGlyLysArg--AlaProSerProGlyAlaTyrLysThrAlaThrProProGlyTyrLys 653
QY 335 CCTCAGCAACCCCTCTGACTTCCGATGGGAGATCACTATCAACTCCAGAGAGGTC 276
DB 653 sProGlySerProProSerPheArgThrGly-----ThrProPro----- 666
QY 275 CTGCATCTCTCTTGAAGGCTTCTCCACTGCTACCTGTCCAGGTGATGTTGG 216
DB 667 -----GlyTyrArgGly-----TherProProAlaGlyProGly 678
QY 215 TGACCAACAA-----GGCCGACAGCTGGGC 191
DB 678 YthrPheLysProGlySerProThrValGly 688

```

## RESULT 6

US-09-046-158A-2  
 Sequence 2, Application US/09046158A  
 Patent No. 6187552  
 GENERAL INFORMATION:  
 APPLICANT: Roberts, Steven L.  
 APPLICANT: Kaytes, Paul S.  
 TITLE OF INVENTION: METHOD FOR IDENTIFYING INHIBITORS OF  
 TITLE OF INVENTION: JAK2/CYTOKINE RECEPTOR BINDING  
 NUMBER OF SEQUENCES: 30  
 CORRESPONDENCE ADDRESS:  
 ADDRESSEE: Pharmacia & Upjohn Co., Intellectual Property  
 ADDRESSEE: Legal Services  
 STREET: 301 Hemlock Street  
 CITY: Kalamazoo  
 STATE: MI  
 COUNTRY: USA  
 ZIP: 49001

COMPUTER READABLE FORM:  
 MEDIUM TYPE: Floppy disk  
 COMPUTER: IBM PC compatible  
 OPERATING SYSTEM: PC-DOS/MS-DOS  
 SOFTWARE: Patent Release #1.0, Version #1.25  
 CURRENT APPLICATION DATA:  
 APPLICATION NUMBER: US/09/046,158A  
 FILING DATE:  
 CLASSIFICATION: 435

ATTORNEY/AGENT INFORMATION:  
 NAME: Darinley Jr., James D.  
 REGISTRATION NUMBER: 33,673  
 TELECOMMUNICATION INFORMATION:  
 TELEPHONE: 616/833-2210  
 TELEFAX: 616/833-8897  
 TELEX: 224401

INFORMATION FOR SEO ID NO: 2:  
 SEQUENCE CHARACTERISTICS:  
 LENGTH: 433 amino acids  
 TYPE: amino acid  
 STRANDEDNESS: single  
 TOPOLOGY: linear  
 MOLECULE TYPE: peptide  
 US-09-046-158A-2

## Alignment Scores:

Pred. No.: 0.00216 Length: 433  
 Score: 116.00 Matches: 103  
 Percent Similarity: 34.30% Conservative: 50  
 Best Local Similarity: 23.09% Mismatches: 149  
 Query Match: 4.53% Indels: 144  
 DB: 4 Gaps: 22

US-09-989-919-15 (1-1397) x US-09-046-158A-2 (1-433)

QY 1206 CAATGAGGAGAAAGGCCCCGCCCCAGCTCTGCG-----TCA 1156  
 : : : : :  
 Db 5 ArgLeuArgArgIstPrpGluGluValIleProAsnProSerIysSerHisLeuPheGln 24  
 : : : : :  
 QY 1155 AACCAACCTGCTCCCTGACCCCAAGTTTGCTCAGTACAGGTGACCTGCCCCCT 1096  
 : : : : :  
 Db 25 AsnGlySerIleAsnLeuTrpProProGlySerMetSerIlePheThrSerGlySerPro 44  
 : : : : :  
 QY 1095 -----GACCCCTGGGCTG-----GGAGTGTCCCTGTC--- 1063  
 : : : : :  
 Db 45 ProHisGlnGlyProItrpGlySerArgPheProGluLeuGluGlyValPheProValGly 64  
 : : : : :  
 QY 1062 ---GGAGAGTCAGAGATAGCTCCCGAGTACAGAAATCACACCAATCTGAGACATCCCC 1006  
 : : : : :  
 Db 65 PheGlyAspSerGluVal-----SerProLeuThrIleGluAspPro 78  
 : : : : :  
 QY 1005 CGCCAAAGTCTCTGTCTGACCTTAAGCTGAGAGAAACATCCACGATGATT---GTCCCC 949  
 : : : : :  
 Db 79 LysHisValCysAspProProSerGlyProAspThrThrProIleAlaSerAspLeuPro 98  
 : : : : :  
 QY 948 AGC-----CAGCCCCACCGCTGCTGTGCGATGTGACGACAGCA 907  
 : : : : :  
 Db 99 ThrGluGlnProProSerProGlnProGlyProProIleAlaSerHisThr-ProGluVal 118  
 : : : : :  
 QY 906 GCAGGGGCTTAACTA-----CATCTCACACTAT 877  
 : : : : :  
 Db 118 ArgIleAlaSerSerPheAspPheAsnGlyProTyrLeuGlyProProHisSerArgSer-L 138  
 : : : : :  
 QY 876 TGAACACAGTATGGGGCTGCCCCAGGACCGTCCAGGGAGGGGACAG-----CCCC 823  
 : : : : :  
 Db 138 eurProAspIleLeuGlyGlnProGluProProGlnGluGlySerGlnLysSerPro 158  
 : : : : :  
 QY 822 CAGTGGGGAGACAGCTGAGCTTGCTTACAGCCCAAGGTTGAGAGAGCACTACATAC 763  
 : : : : :  
 Db 158 robProGlySerLeuGluTyrLeuCysLeu---ProAlaGlyGlnValGlnLeuValP 177  
 : : : : :  
 QY 762 CCCATGATTCCTGAGTACAGGACGAGCAGATCCATAGCTAGCTGTAACATCATTTT 703  
 : : : : :  
 Db 177 ro----- 177  
 : : : : :  
 QY 702 AACATTGGCATTAATGACAGCTTTGCTGTGTCATCTGTCTGGAGTTAGCTT 643  
 : : : : :  
 Db 178 -----LeuAlaG 180  
 : : : : :  
 QY 642 TGTCTGAGGGGTTCCAGGTGACATGACATGCTCCAGTGCACATCTCCTCCACAG 583  
 : : : : :  
 Db 180 ImlaMetGlyProGlyGln-----AlaValGluValGluArgArgProSerGln 197  
 : : : : :  
 QY 582 GCAAGCCACAGCTGTGATAGTACAGTGTGACGCGAGCCCTCAGATAGTGCAGCAG 523  
 : : : : :  
 Db 197 ly-----AlaIleAspIleSerProSerLeuGluSerGlyGly--- 208  
 : : : : :  
 QY 522 AGGCTTCACATATGTGCTTGTGTTGGATCAGAGGACCACTTGCTTCATCCATCTTGA 463  
 : : : : :  
 Db 209 -----GlyProAlaProProAlaLeuGlyPro----- 217  
 : : : : :  
 QY 462 AAAAGACAGCTCCCACTGGGGGGAGAGGTGAGAGTAAACAGGGCTGAGTGTTC 403  
 : : : : :  
 Db 218 -----ArgValGlyGlyGlnAspGlnLysAspSerProValAlaIlePro 233  
 : : : : :  
 QY 402 TTCTTGAGGAGCAAGCATTTCTCCCTTAATTGACTGTGCTGAAGACACGATACCTCTCT 343  
 : : : : :  
 Db 233 etSerSerGlyAspThrGluAspProGlyValAlaSerGlyTyrValSer-SerAlaAsp 252  
 : : : : :  
 QY 342 CTGATCACTCAGACACCCCTCTGACTTCGATGGGGAGTACATATCTATCACTC--- 286  
 : : : : :  
 Db 253 LeuVal-----PheThrProAsnSerGlyAlaSerSerValSerLeuVal 267  
 : : : : :  
 QY 285 -----CAGAGAGTCTGCTGCACTCTCCCTTGAGAGGGCTTCT--- 250  
 : : : : :  
 Db 268 ProSerLeuGlyLeuProSerAspGlnThrProSerLeuGlyLeuAlaSerGly 287

QY 249 CCCACTGGCTACCTGTCCAGTGT-----GTCTGTGTGTGAC 211  
 : : : : :  
 Db 288 ProProGlyAlaIleProIleProValLysSerGlyPheGluGlyTyrValGluLeuProPro 307  
 : : : : :  
 QY 210 ACAAGAGCCGACACTGGGATGCTCTCACAGACATCCACAGCT--- 165  
 : : : : :  
 Db 308 IleGluGlyArg-SerProArgSerProArgAsnProValProProGluAlaLysSe 327  
 : : : : :  
 QY 164 -----CAGCCAGGTTGAACACTGAAGAGGACAGCTCCGCTGTGTGATGAGC 115  
 : : : : :  
 Db 327 rProValLeuAsnProGly-----GluArgProAlaAspValSe 340  
 : : : : :  
 QY 114 CAGACAGGTAGTCTTCTGGGGATGAGTGTCTGTGAGTACATCTGTAATG 55  
 : : : : :  
 Db 340 rProThrSerProGlnProGluGlyLeuLeuValLeuGlnValGlyAsp---TyrCy 359  
 : : : : :  
 QY 54 CTGCTTGGCG 45  
 : : : : :  
 Db 359 spHeuPro 362

## RESULT 7

US-08-783-774-2

Sequence 2, Application US/08783774

Patent No. 6054130

GENERAL INFORMATION:

APPLICANT: Spaete, Richard

TITLE OF INVENTION: NON-SPLICING VARIANTS OF

NUMBER OF SEQUENCES: 19

CORRESPONDENCE ADDRESS:

ADDRESSER: Pennie &amp; Edmonds

STREET: 1155 Avenue of the Americas

CITY: New York

STATE: NY

COUNTRY: USA

ZIP: 10036/2711

COMPUTER READABLE FORM:

MEDIUM TYPE: Diskette

COMPUTER: IBM Compatible

OPERATING SYSTEM: DOS

SOFTWARE: FastSeq Version 2.0

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/08/783,774

FILING DATE: 15-JAN-1997

CLASSIFICATION: 435

ATTORNEY/AGENT INFORMATION:

NAME: Coruzzi, Laura A.

REGISTRATION NUMBER: 30,742

REFERENCE/DOCKET NUMBER: 7682-037

TELEPHONE: 212-790-9090

TELEFAX: 212-869-8864

TELEX: 66141 PENNIE

INFORMATION FOR SEQ ID NO: 2:

SEQUENCE CHARACTERISTICS:

LENGTH: 907 amino acids

TYPE: amino acid

STRANDEDNESS:

TOPOLOGY: unknown

MOLECULE TYPE: protein

US-08-783-774-2

## Alignment Scores:

Pred. No.: 0.00378

Score: 115.00

Percent Similarity: 32.77%

Best Local Similarity: 21.69%

Query Match: 4.49%

DB: 3

Length: 907

Matches: 90

Conservative: 46

Mismatches: 117

Indels: 162

Gaps: 18

US-09-989-919-15 (1-1397) x US-08-783-774-2 (1-907)

```

1182 CCCACCCCGCCAGT---CCTGGCTCAACACAC--- 1147
1146 TGCTCCCTGACCCAGCTTGCTGAGTGTGACCTGCTGCTGACCCCTG 1087
462 AlaserThrclyProthrValSerThrAlaAspVal-----TherPro----- 476
1086 GCTGGCTGGAGAGTCCCTGCGGAGGTGAGATAGCTCCCGACGTACAGATCA 1027
477 -----ThrProAlaGlyThrThrseryAlaSerProValThrProSer 491
1026 CCCACATCTGAGCATCCCGCCCAAGTCTCTGTCACACCTTAAGCTGAGAACACT 967
492 ProSerProThrAspAsnGlyThrGlySerIyAlaProAspMetThrSerThrSer 511
966 CCAGCATGATGTGCTCCCGACGAGCCCGCTGCTGGCATGTGACAGACAGCA 907
512 Pro-----ValThrThrProThrProAsnAlaThrSerProThrProAlaVal 527
906 GCAGGGCTTTAGCTACATTCTACACTATTGAACAGTATGAGGGCTGCCACGGAAC 847
528 Thr----- 528
846 GCTCCAGGAGGGGACAGCCCGCCAGTGGGGAGACTAGCCTGAGCTTGCTTACAGCCCA 787
529 -----ThrPro 530
786 CAGGGTTGAGAGGAGCATACCCCATGATT----- 754
531 ThrProAsnAlaThrSerProThrProAlaValThrThrProThrProAsnAlaThrSer 550
753 CCTGAGTAGAGAGCAGACAGTCCCATAGCTAGGCTGTAACATCAATTTAACTTTGG 694
551 ProThrLeuGlyLysThrSerProThrSerAlaValThrThrProThrProAsn----- 568
693 CATATTGACAGTTTGCTGCTGTCACATCTGCTGCTGGAGTTCCTTGTCTGAC 634
569 -----AlaThrSerProThrLeuGlyLysThr-----SerPro 579
633 GGGTTCCAGGTAGCATGACAGTGTGCGCAGTGCAGTCACTCCCTCCACAGCAG--- 578
579 othrSerAlaValThrThrProThrProAsnAlaThr-SerProThrLeuGlyLysThrS 599
577 -----CCAGCTGCTGAGGATGACTGCTGCTGAC 547
599 erProThrSerAlaValThrThrProThrProAsnAlaThrGlyProThrValGlyLut 619
546 CGAGCCTCAG----- 536
619 hrSerProGlnAlaAsnAlaThrAsnAlaThrLeuGlyLysThrSerProThrProVal 639
535 --ATAGGTGCGCAGAGCCCTTGCATATGTGCTTTGGATCAGGACCACTGGC 478
639 alThrSerGlnProLysAsnAlaThrSerAlaVal-----ThrThrGly 654
477 TCCATTCAGCTTTAAAAAGAC-----AGTGCAGACTGGGGGGGAGAGGT 430
654 lnhIsaenIleThrSerSerThrSerSerMetSerLeuArgProSer----- 670
429 GAGAGGTAAAGGGCTGATGTTCTCTTTTGAGAGCAGACCACTTCCTCAATTGA 370
671 -----SerAsnProGlnThrThrLeuSerProSerThrS 681
369 CTGCTTGAAGACAGCATACCTCTCTCATACCTCAGACCCCTGACTTCGCAT 310
681 erAspAsnSerThrSer-HisMetProLeuLeuThrSerAlaHisPro-----Thr 697
309 GGGGATCACTATTTATCACTCCAGAGAGTCTGACATCTCCCTGAGAGGCTTCT 250
698 GlyGlyGlu-----AsnIle-ThrGlnValThrProAlaSerIle----- 710

```

```

QY 249 CCCACTGCTACCTGTCCAGGTGTGTGTGTGACCAAGCCCGACACTGGGCA 190
DB 711 -----SerThrHisH 714
QY 189 TGGCTTCACAGACATTCACAGCTTCCAGCTTGAACACTGAAAGAGGAGCTCCG 130
DB 714 sValSerThrSerSerProGlnProArgProGlyThrThr-----SerGlnAlaSer 731
QY 129 TGTGTGTAGAGTGGCCAGCGGCTAGCTTCCTGGG 93
DB 731 rGlyProGlnAsnSerSerThrThrLysProGly 743

RESULT 8
US-09-328-599A-1
; Sequence 1, Application US/09328599A
; Patent No. 6432679
; GENERAL INFORMATION:
; APPLICANT: MOND, James J. and Lees, Andrew
; TITLE OF INVENTION: Enhancement of B Cell Activation by
; TITLE OF INVENTION: Co-ligation of Receptors for Antigen and Complement C3d
; TITLE OF INVENTION: Using EBV gp350/220 or EBV gp350/220 Peptide Adjuvants
; NUMBER OF SEQUENCES: 2
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Finnegan, Henderson, Farabow, Garrett &
; ADDRESS: 1300 I Street, N.W., Suite 700
; CITY: Washington
; STATE: DC
; COUNTRY: USA
; ZIP: 20005
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; OPERATING SYSTEM: IBM PC compatible
; SOFTWARE: Patent in Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/328,599A
; FILING DATE:
; CLASSIFICATION:
; ATTORNEY/AGENT INFORMATION:
; NAME: Fordis, Jean B.
; REGISTRATION NUMBER: 32,984
; REFERENCE/DOCKET NUMBER: 04995, 6025-00000
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (202)408-4000
; TELEFAX: (202)408-4400
; INFORMATION FOR SEQ ID NO: 1:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 907 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULAR TYPE: protein
; US-09-328-599A-1

Alignment Scores:
Pred. No.: 0.00378 Length: 907
Score: 115.00 Matches: 90
Percent Similarity: 32.77% Conservative: 46
Best Local Similarity: 21.69% Mismatches: 117
Query Match: 4.49% Indels: 162
DB: 4 Gaps: 18

US-09-989-919-15 (1-1397) x US-09-328-599A-1 (1-907)
QY 1182 CCCACCCCGCCAGT---CCTGGCTCAACACAC--- 1147
DB 442 ProAsnThrThrThrclyLeuProSerSerThrHisValProThrAsnLeuThrAlaPro 461
QY 1146 TGCTCCCTGACCCAGCTTGCTGAGTGTGACCTGCTGCTGACCCCTG 1087
DB 462 AlaserThrclyProthrValSerThrAlaAspVal-----TherPro----- 476

```

OY	1086	GCTGGCCTGGAGAGTGTCCCTGTCGGAGAGGTGAGATAGATGCTCCCGACGGTGAACAATCA	1022
Db	477	-----ThrProAlaGlyThrThrSerGlyAlaSerProValThrProSer	491
OY	1026	CCCAATCCTGGAGAGCATCCCGCCGCAAGATCTCTGTCGACAGACTTAAGCTGAGAGAACT	967
Db	492	ProSerProTPrAspAsnGlyThrIuSerLysAlaProAspMetThrSerSerThrSer	511
OY	966	CCAGCATGAGATTTGTCGCCAGCAGGCCCCACGCTGCTGGTGAGCATGTGCAGACAGACA	907
Db	512	Pro-----ValThrThrProThrProAsnAlaThrSerProThrProAlaVal	527
OY	906	GCAGGGGCTTTAGTACATTTCTCAACATTTGAACAAGTATGGGGCTGCCACAGAAC	847
Db	528	Thr-----	528
OY	846	CGTCCAGAGAGGGGACAGACCCCGAGTGGGAGACTAGACCTGAGCTTTACAGCCCA	787
Db	529	-----ThrPro	530
OY	786	CAGGTTGAGAGGAGCATATCCCATGATT	754
Db	531	ThrProAsnAlaThrSerProThrProAlaValThrThrProThrProAsnAlaThrSer	550
OY	753	CCTGAGCATAGAGCCAGCAGTCCCATCTAGCTAGCGTGAACATCAATTTAATCATTTGG	694
Db	551	ProThrLeuGlyLysThrSerProThrSerAlaValThrThrProThrProAsn-----	568
OY	693	CATTATTCACAGTTGTGCTGTGCATCTGTCTCTGGAGTGTAGCTTTGTCTGCAG	634
Db	569	-----AlaThrSerProThrLeuGly-LysThr-----SerPr	579
OY	633	GGGTTCCAGGTGACATGATGATGCTGTCAGATGCAAGTCACTCCCTCCACAGGCAAG---	578
Db	579	oThrSerAlaValThrThrProThrProAsnAlaThr-SerProThrLeuGlyLysThrS	599
OY	577	-----CCAGCTGCTGAGGATGATGACGTGTCAGC	547
Db	599	erProThrSerAlaValThrThrProThrProAsnAlaThrGlyProThrValGlyLut	619
OY	546	CGAGCCCTCAG-----	536
Db	619	hrSerProGlnAlaAsnAlaThrAsnAlaThrLeuGlyGlyThrSerProThrProVal	639
OY	535	--ATAGTTCAGCAGAGGCTTCAATATGTGTTGTTGGATCAGAGGACCACTTGGC	478
Db	639	alThrSerGlnProLysAsnAlaThrSerAlaVal-----ThrThrGlyG	654
OY	477	TCCATCCAGTCTTGA AAAAGACC-----AGCTGCCAGCTGGGGGGCAGAGGT	430
Db	654	lnHisAsnAlaThrSerSerSerThrThrSerMetSerLeuLysProSer-----	670
OY	429	GAGAGGTAAAGGGGTGATGTTTCTTTCTTGAGGGCAAGACATTCCTCCATATTGA	370
Db	671	-----SerAsnProGlnThrLeuSerProSerThrS	681
OY	369	CTGCTTGAAGACAGGTATACCTCTCTGCATCACTCAAGCAACCCCTCTGACTTCCGAT	310
Db	681	erAspAsnSerThrSer-HisMetProLeuLeuThrSerAlaHisPro-----Thr	697
OY	309	GGGGATCACTATATATCAACTCCACAGAGGTCTGCCATGTCCCTTTGAAAGGGCTTCT	250
Db	698	GlyGlyGlu-----AsnLeu-ThrGlnAlaThrProAlaSerIle-----	710
OY	249	CCCACTGGCTACCTGTCCAGTGGTCTGGTGGAGCAACAAGGCGCAACTGGGCA	190
Db	711	-----SerThrHisHis	714
OY	189	TGGCTTCACAGATCCACAGCCTTCAAGCAGGTTGAACACTGAAGAAGGAGCAGCTCCG	130
Db	714	sValSerThrThrSerSerProGlnProLysGlyProLysThr-----SerGlnAlaSer	731
OY	129	TGATGTAGATGAGCCAGCAGCGGATGATCTTCTCTGGG	93

```

Db          731 rGlyPrdGlyAsnSerSerThrSerThrLysProGly 743
RESULT 9
PCT-US95-04611A-19
: Sequence 19, Application PC/TUS9504611A
: GENERAL INFORMATION:
: APPLICANT: Spaete, Richard and Jackman, Winthrop, T.
: TITLE OF INVENTION: Non Splicing Variants of gp350/220
: NUMBER OF SEQUENCES: 19
: CORRESPONDENCE ADDRESS:
: ADDRESSEE: Cooley Godward Castro Huddleston & Tatum
: STREET: 5 Palo Alto Square
: CITY: Palo Alto
: STATE: California
: COUNTRY: USA
: ZIP: 94306
: COMPUTER READABLE FORM:
: MEDIUM TYPE: Floppy disk
: COMPUTER: IBM PC compatible
: OPERATING SYSTEM: PC-DOS/MS-DOS
: SOFTWARE: Patent in Release #1.0, Version #1.25
: CURRENT APPLICATION DATA:
: APPLICATION NUMBER: PCT/US95/04611A
: FILING DATE:
: CLASSIFICATION:
: PRIOR APPLICATION DATA:
: APPLICATION NUMBER: 08/229,291
: FILING DATE: April 18, 1994
: CLASSIFICATION:
: ATTORNEY/AGENT INFORMATION:
: NAME: Luann Ceert
: REGISTRATION NUMBER: 31,822
: REFERENCE/DOCKET NUMBER: AVIR-003/00US
: TELECOMMUNICATION INFORMATION:
: TELEPHONE: 415-843-5163
: TELEFAX: 415-857-0663
: TELEX: 380816 CooleyPA
: INFORMATION FOR SEQ ID NO: 19:
: SEQUENCE CHARACTERISTICS:
: LENGTH: 907 amino acids
: TYPE: amino acid
: TOPOLOGY: linear
: MOLECULE TYPE: protein
: PCT-US95-04611A-19

Alignment Scores:
Pred. No.: 0.00378 Length: 907
Score: 115.00 Matches: 90
Best Local Similarity: 32.77% Conservative: 46
Best Local Similarity: 21.69% Mismatches: 117
Query Match: 4.49% Indels: 162
DB: 5 Gaps: 18

US-09-989-919-15 (1-1397) x PCT-US95-04611A-19 (1-907)
QY 1182 CCACCCCCGCCCGCCAGGT--CCTGGCTCAACACGACC----- 1147
||| ||| ||| ||| |||
Db 442 ProasnthrThrThnglyLeuProSerSerThrHisValProThrAsnLeuThrAlaPro 461
||| ||| ||| ||| |||
QY 1146 TGCTTCCTGACCCCGACGCTTGAGCTCAGTCAGTAGGTGACCTGCGCCCTGACCCCTGG 1087
||| ||| ||| ||| |||
Db 462 AlaSerThrnglyProThrValSerThrAlaAspVal-----ThnSerPro----- 476
||| ||| ||| ||| |||
QY 1086 GCTGGCCTGGAGGTGCTCCCTGTGTGGGAGGTCAGAGATAGCCTCCCGAGTACAGATCA 1022
: : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 477 -----ThnProAlaGlyThrThrSerGlyAlaSerProValThnProSer 491
: : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 1026 CCCACATCTGAGAGATCCCGCCGACAGTCTCCTGTCCGACGACTTAAGCTGAGAGACA 967
||| ||| ||| ||| |||
Db 492 ProSerProTrpAspAsnGlyThrGluSerLysAlaProAspMetThrSerSerThrSer 511
||| ||| ||| ||| |||
QY 966 CCAGATGATTTGTCCACGACGACGACCCGACGCGCTGTGTGGCATGTGACGACGACGA 907

```



```

Db      512 Pro-----ValThrThrProThrProAsnAlaThrSerProThrProAlaVal 527
QY      906 GCGGGGGCTTACTTACTTCTACACTATTGAACAGATGGGGCTCCACGGAAC 847
Db      528 Thr-----
QY      846 CGTCCAGAGGAGGAGCAGCCCCAGTGGGAGACTAGCCTGAGCTTGCTTACAGCCCA 787
Db      529 -----ThrPro 530
QY      786 CAGGGTTGAGAGCAGTACATACCCCATGATT----- 754
Db      531 ThrProAsnAlaThrSerProThrProAlaValThrThrProThrProAsnAlaThrSer 550
QY      753 CCGGAGTACAGAGCAGCAGATCCCATATGACTAGGTGTAACTACATTTTACATTGG 694
Db      551 ProThrLeuGlyThrSerProThrSerAlaValThrThrProThrProAsn----- 568
QY      693 CATTATTGACGTTTGCTGCTGACATCTGCTGCTGGATGTTAGCTTGTCTGACAG 634
Db      569 -----AlaThrSerProThrLeuGly-LysThr-----SerPr 579
QY      633 GGGTCCAGAGTACATGACATGCTGCTGCTGACATGACATCTCCCTCCACAGGAG----- 578
Db      579 oThrSerAlaValThrThrProThrProAsnAlaThr-SerProThrLeuGlyLysThrS 599
QY      577 -----CCAGCTCTAGAGATAGTACGCTGCTGACG 547
Db      599 erProThrSerAlaValThrThrProThrProAsnAlaThrGlyProThrValGlyGln 619
QY      546 CGAGCCCTGAG----- 536
Db      619 hrSerProGlnAlaAsnAlaThrAsnHisThrLeuGlyLysThrSerProThrProVal 639
QY      535 --ATAGGTACAGCAGAGGCTTCTTCATATGTTGTTGGATCAGGACCATTTGGC 478
Db      639 alThrSerGlnPolyAsnAlaThrSerAlaVal-----ThrThrGly 654
QY      477 TCATCCAGCTTGAATAAAC-----AGCTCCGACCTGGGGGAGAGGT 430
Db      654 lnhIsnAlaThrSerSerSerSerSerSerSerSerSerSerSerSerSerSerSer 670
QY      429 GAGAGGTACAGAGGCTGATGTTCTCTTCTGAGAGCAGACCATCTCCCTAATTGA 370
Db      671 -----SerAsnProGlnThrLeuSerProSerThrS 681
QY      369 CTGCTTGAAGACAGTATACCTCTCTCATCATCACTCAGACCCCTCTGACTTCGAT 310
Db      681 erAspAsnSerThrSer-HisMetProLeuLeuThrSerAlaHisPro-----Thr 697
QY      309 GGGGATCACTATCTATCAACCTCCAGAGGCTCTCCATCTCCCTTGAGAGGGCTTCT 250
Db      698 GlyGlyGly-----AsnIle-ThrGlnValThrProAlaSerIle----- 710
QY      249 CCACATGGCTCACCTGTCAGAGTGTGTGTTGGTACCAACAAGCCGACACATGGGCA 190
Db      711 -----SerThrHisIle 714
QY      189 TGGCTCTACAGACATCCACAGCCTCAGCAGGTTGAACACTGAAGAGAGCAGCTCCG 130
Db      714 sValSerThrSerSerProGlnProArgProGlyThrThr-----SerGlnAla 731
QY      129 TGGTGTGATGAGCCAGCAGCGGTAGCTTCTCTGG 93
Db      731 rGlyProGlyAsnSerSerThrSerThrLysProGly 743

```

RESULT 10  
 US-08-127-499A-8  
 Sequence 8, Application US/08127499A  
 Patent No. 5510264  
 GENERAL INFORMATION:  
 APPLICANT: VAN ALSTYNE, Diane

```

APPLICANT: SHARMA, Lawrence Rajendra
TITLE OF INVENTION: ANTIBODIES WHICH BIND MENINGITIS RELATED
TITLE OF INVENTION: HOMOLOGOUS ANTIGENIC SEQUENCES
NUMBER OF SEQUENCES: 40
CORRESPONDENCE ADDRESS:
ADDRESSER: Foley & Lardner
STREET: 3000 K Street, N.W., Suite 500
CITY: Washington
STATE: D.C.
COUNTRY: USA
ZIP: 20007-5109
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
FILING DATE: 28-SEP-1993
APPLICATION NUMBER: US/08/127,499A
ATTORNEY/AGENT INFORMATION:
NAME: BENT, Stephen A.
REGISTRATION NUMBER: 29,768
REFERENCE/DOCKET NUMBER: 51916/102/INBI
TELECOMMUNICATION INFORMATION:
TELEPHONE: (202) 672-5300
TELEFAX: (202) 672-5399
TELEX: 904136
INFORMATION FOR SEQ ID NO: 8:
SEQUENCE CHARACTERISTICS:
LENGTH: 1063 amino acids
TYPE: amino acid
STRANDEDNESS:
TOPOLOGY: unknown
US-08-127-499A-8
Alignment Scores:
Pred. No.: 0.00406 Length: 1063
Score: 115.00 Matches: 95
Percent Similarity: 31.65% Conservative: 37
Best Local Similarity: 22.78% Mismatches: 116
Query Match: 4.49% Indels: 149
DB: 1 Gaps: 20
US-09-989-919-15 (1-1397) x US-08-127-499A-8 (1-1063)
QY      1184 GCCCCACCCCCCGCCAGTCTGCTCAAC-----ACACTCTCTCCCTGACCC 1134
Db      77 AlaProProProProGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGln 96
QY      1133 CAGTCTGCTCATGTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1074
Db      97 SerArg-----AlaPro-----ProGlnGlnProGln----- 105
QY      1073 TGTCCCTGTCGGAGGTGAGATAGCTCCCTCCAGTACAG-----AAT 1029
Db      106 -----ProArgArgMetGlnThrGlyArgGlyGlySer 116
QY      1028 CACCCATCTCTGAGATCCCGCCCAAGTCTCTGCTCAGACCTTAAGCTGAGAGA----- 972
Db      117 AlaProArgProGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGln 136
QY      971 -----ACACTCCAGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 926
Db      137 LeuArgProProLeuHisAspPro-----AspThrGln-AlaProThrGlnAlaCysVal 155
QY      925 -----GGCATGTCAGAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAG 870
Db      155 rSerThrLeuThrProLeuGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGln 175
QY      869 AGTATGGGGCTGCCACGGAACCGTCCAGGCA----- 836
Db      175 rAsnLeuGlyThrProProLeuAspGlnAspGlyArgTrpAspProAlaLeuMetLys 195

```



```

Db      214 yAspValArgGlyValThrGlyLysGlyGluArgThrTyraAlaGluLysAspPheArgVa 234
QY      722 GGCTGGTAAATCACTATTAACTTTGGCATTTCACAGCTTTGCTGCTGACATCTG 663
Db      234 LgLyGly-----ThraGTrpHisArgLeuLeu-----ArgMetPr 246
QY      662 TCGTGTGGAGTGTAGCTTTGCTGTGACAGGGGTCCAGAGTACATGACAGTGGCCAGT 603
Db      246 oValArgGlyLeuAspGlyAspSerAlaProLeuProPro---HisThrThrGluAlaGly 265
QY      602 GCAAGTCACTCCCTCCACAGGCAAGCCAGCTGCTGAGATGACAGTGGTCAAGCCG-- 545
Db      265 eGluThrArg-----SerAlaArgHisProThr 274
QY      544 -AGCCCTCAATAGTGTACAGCCAGAGGCTTCACATATGTGTCTTTGGTGGATCAGGAC 486
Db      274 pArgLLeuArgPheGlyAlaProGlnAlaPheLeuAlaGlyLeuLeuLeuAlaThrValAl 294
QY      485 CACTGTGCTCCATCCAGCTTTGAAAGAACAGCAGCTGCGGAGCTGGGGGAGAGTGAGA 426
Db      294 a-ValGlyThrAlaArgAlaGlyLeuGlnProAlaGlnAspMet----- 308
QY      425 GGTAAACAGGCTGATGTTTCTCTTTCTGAGGCAAGACCATTCCTCCATATTGACTGC 366
Db      308 ----- 368
QY      365 CTGGAAGACAGTATACCTCTCTCTCATCACTGACACCCCTCTGATTCGGATGGG 306
Db      309 -----AlaAlaProProThr-LeuProGlnPro 317
QY      305 GATCACTATCTATCAACCTCCAGAGAGTCTCGCATCTCCCTGAAAGGCTTCTCCCA 246
Db      318 -----Pro 318
QY      245 CTGGCTCACTGTCCAGAGTGTCTGTGTGTGACCAAGGCCCACTGGGATGGC 186
Db      319 CysAlaHisGlyGlnHisTyrgLysHisHisHisGlnLeuProPheLeuGlyHisAsp 338
QY      185 TCTCCACAGATCCACAGCTCCAGGCTGAAACAGTGAAGAGGACACTCCCGTGT 126
Db      339 GlyHisHis-----GlyGlyThrLeuArgVal 347
QY      125 GGTAGATGCGCAGACAGCGTAGT-----CTTCTGGGGGATGG 87
Db      348 GlyGlnHisTyrgAlaGlnAlaSerAspValLeuProGlyHisTyrLeuGlnGlyGlyTyr 367
QY      86 TGCTGTCTGGGA-----TACACTGTACTGGTACTGC 54
Db      368 GlyCysTyraAsnLeuSerAspTrpHisGlnGlyThrHisValCys 382

```

```

RESULT 12
US-08-874-569B-20
; Sequence 20, Application US/08874569B
; Patent No. 6306650
; GENERAL INFORMATION:
; APPLICANT: Townes, Tim M.
; APPLICANT: Donze, David
; TITLE OF INVENTION: DELTA-BRYTHROID KRUPPEL-LIKE FACTORS AND
; TITLE OF INVENTION: METHODS OF USE
; FILE REFERENCE: 05118.000802
; CURRENT APPLICATION NUMBER: US/08/874,569B
; CURRENT FILING DATE: 1997-06-13
; PRIOR APPLICATION NUMBER: 60/019,769
; PRIOR FILING DATE: 1996-06-14
; NUMBER OF SEQ ID NOS: 21
; SOFTWARE: FastSeq for Windows Version 4.0.
; SEQ ID NO 20
; LENGTH: 376
; TYPE: PRT
; ORGANISM: Murine Beta-EKLF
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence:\ No. 6306650e =
; OTHER INFORMATION: synthetic construct

```

```

US-08-874-569B-20
Alignment Scores:
Pred. No.: 0.00464 Length: 376
Score: 112.50 Matches: 80
Percent Similarity: 32.23% Conservative: 27
Best Local Similarity: 24.10% Mismatches: 97
Query Match: 4.39% Indels: 128
DB: 4 Gaps: 18

```

US-09-989-919-15 (1-1397) x US-08-874-569B-20 (1-376)

```

QY      1179 ACCCCGCGCCAGGT-----CTGGCTAAACACACCTGCTCCGACCCAGTCC 1129
Db      98 ThrAspPheProGlySerGluSerProGlyThrSerArgThrCysAlaLeuAlaProSer 117
QY      1128 TTGGCTCATGCTGAGGTGTGACCTTGTGCGCCCTGACCCCTGGGCTGGGAGTGTC 1069
Db      118 ValGlyPro-----ValAlaGlnPheGluProProGluSerLeuGlyAlaTyr 133
QY      1068 CCTGTGGAGGTGACAGATAGCTTCCCA---GGTACAGATACCCACATCTGGAGC 1012
Db      134 AlaGlyGlyProGlyLeuValThrGlyProLeuGlySerGlnGlyHisThrSerTrpAla 153
QY      1011 ATCCCGCGCAAGTCTCTGTCCAGACTTACAGTGAAGAACATCTCAGATGATGTC 952
Db      154 HisPro-----ThrProArg----- 158
QY      951 CCCACAGACCCCGCCAGCTGTGTGATGTGACAGACAGACAGAGGGCTTAACT 892
Db      159 -----ProProAlaProGluProPheValAlaProAlaLeuAlaProGlyLeuAla 175
QY      891 ACATTCTACACTATTGAACACAGTATGGGCTGCCACGAGAACCGTCCAGG----- 838
Db      176 -----ProLysAlaGlnProSerTyrSer 183
QY      837 -----GAGGGACAGAGCCCGGAGGGA----- 814
Db      184 AspSerArgAlaGlySerValGlyGlyPhePheProArgAlaGlyLeuAlaValProAla 203
QY      813 -----GACTAGCTGAGCTTGTAAAGCCACAG 784
Db      204 AlaProGlyAlaProGlyLeuLeuSerGlyTyrProAlaLeuTyrProAlaProGln 223
QY      783 GGTGAGAGGACAGTATACCCCATGATCTCTGACTGAGACAGACAGTCCATAGCT 724
Db      224 ---TyrGlnGly---HisPheGlnLeuPheArgGlyLeuAlaAlaProSerAlaGlyGly 241
QY      723 AGGCTGTAACTCACTATTAACTTTGACATTATGACAGTTTCTGTCGTCATCTG 664
Db      242 ThrAlaProProSerPheLeuAsn----- 249
QY      663 GTCTGTCTGGATGTAGTTAGTTGTCTGACAGGGTTCCAGGTGACATGACAGTCTCCAG 604
Db      250 ---CysLeuGly-----ProGly-ThrValAlaThrGluLe 260
QY      603 TGAAGTCACTCCCTCCACAGGCAAGCCAGCTGCTGAGAGATAGTACGTGTCACCCA 544
Db      260 uGlyAlaThrAlaAlaLeuAlaGlyAspAlaGlyLeu-----SerProGlyThrAl 276
QY      543 GCCCTCAGATAGTGTGACAGAGGCTT-----CACATATGTGG 505
Db      276 aProProLysArgSerArgArgThrLeuAlaProLysArgGlnAlaHisThrCysG 296
QY      504 TCTTGTGGG-----ATCAGGACCACTTGCTCC 475
Db      296 yHisGlnGlyCysGlyLysSerTyrSerLysSerHisLeuLysAlaHisLeuArgTh 316
QY      474 ATCCAGTCTTGAAGAACACAGCTGCCAGCTGG---GGGACAGAGTGA 426
Db      316 HisThrGlyGlyLysProTyrAlaCysSerTrpAspGlyCysAspTrpArgPheAlaAr 336
QY      425 -----GTTAACAGGCTGAGATGTTCT-- 405

```

Db 336 gSerAspGluLeuThrArgHisTyrArgLysHisThrGlyHisArgProPheCysCysG1 356  
 QY 404 -----TCTTTCGAGAGCAACCAT 384  
 Db 356 yLeuCySProArgAlaPheSerArgSerAspHis 367

RESULT 13  
 US-08-127-499A-1  
 Sequence 1, Application US/08127499A  
 Patent No. 5510264  
 GENERAL INFORMATION:  
 APPLICANT: VAN ALSTYNE, Diane  
 APPLICANT: SHARMA, Lawrence Rajendra  
 TITLE OF INVENTION: ANTIBODIES WHICH BIND MENINGITIS RELATED  
 TITLE OF INVENTION: HOMOLOGOUS ANTIGENIC SEQUENCES  
 NUMBER OF SEQUENCES: 40  
 CORRESPONDENCE ADDRESS:  
 ADDRESSEE: Foley & Lardner  
 STREET: 3000 K Street, N.W., Suite 500  
 CITY: Washington  
 STATE: D.C.  
 COUNTRY: USA  
 ZIP: 20007-5109  
 COMPUTER READABLE FORM:  
 MEDIUM TYPE: Floppy disk  
 COMPUTER: IBM PC compatible  
 OPERATING SYSTEM: PC-DOS/MS-DOS  
 SOFTWARE: Patent In Release #1.0, Version #1.30  
 CURRENT APPLICATION DATA:  
 APPLICATION NUMBER: US/08/127,499A  
 FILING DATE: 28-SEP-1993  
 ATTORNEY/AGENT INFORMATION:  
 NAME: BENT, Stephen A.  
 REGISTRATION NUMBER: 29,768  
 REFERENCE/DOCKET NUMBER: 51916/102/INBI  
 TELECOMMUNICATION INFORMATION:  
 TELEPHONE: (202)672-5300  
 TELEFAX: (202)672-5399  
 TELEX: 904136  
 INFORMATION FOR SRO ID NO: 1:  
 SEQUENCE CHARACTERISTICS:  
 LENGTH: 992 amino acids  
 TYPE: amino acid  
 STRANDEDNESS:  
 TOPOLOGY: unknown  
 US-08-127-499A-1

Alignment Scores:  
 Pred. No.: 0.008 Length: 992  
 Score: 112.00 Matches: 97  
 Percent Similarity: 31.9% Conservative: 38  
 Best Local Similarity: 22.9% Mismatches: 129  
 Query Match: 4.38% Indels: 159  
 DB: 1 Gaps: 22

US-09-989-919-15 (1-1397) x US-08-127-499A-1 (1-992)

QY 1184 GCGCCACCCCGCCGCGCTGCTGCTCAACC-----ACACTGCTCCCTGACCC 1134  
 Db 76 AlaProProProProGluArgGlnGlnSerArgSerGlnThrProAlaProLysPro 95  
 QY 1133 CAGCTTGAGCTATGCTGAGAGTGTGACCTGTGCGCCCTGACCCCTGGCGCTGGCGTGGAG 1074  
 Db 96 SerArg-----AlaPro-----ProGlnGlnProGln-----104  
 QY 1073 TGTCCCTGTGCGGAGGTGAGATAGCTCCCGAGGTACAG-----AAT 1029  
 Db 105 -----ProProArgGlnGlnThrGlnArgGlnGlnSer 115  
 QY 1028 CACCCACATCTGTGAGATCCCGCCCAAGTCTCTCCAGACCTAGAGTGAGAGA--- 972  
 Db 116 AlaProArgProGluLeuGlyProProThrAsnProPheGlnAlaAlaValAlaArgGly 135

QY 971 -----ACATCCAGATGATTTGCCACAGCGCCCGCCGCTGCTG----- 926  
 Db 136 LeuArgProProLeuHisAspPro-----AspThrGln-AlaProThrGlnAlaCysValTh 154  
 QY 925 ---GGCATGTGACGACGACGACGAGGCGCTTTAGCTACATTTGACATTTAGAAC 870  
 Db 154 rSerThrLeuTrpSerGlnGlyGlnGlyAlaValPheTyrArgValAspLeuHisPhe11 174  
 QY 869 AGTGAATGGGGCTGCGCCAGGGAACCGTCCAGGGA----- 836  
 Db 174 eAsnLeuGlyThrProProLeuAspGluAspGlyArgTyrAspProAlaLeuMetTyrAs 194  
 QY 835 -----GGGACACAGCCCGCCAGTGGGAGAGACTAGGCTGCTGCTTACGCCACAG 783  
 Db 194 nProCysGlyProGluProPro-----AlaHisValAlaValAlaTyrAsnGlnProAlaG1 213  
 QY 782 GTTGAGAGGACAGTATACCCCGATTTCTGAGCTAGAGGACAGCAGTCCCATAGCTA 723  
 Db 213 yAspValArg-----GlyValTrpGlyGlyGlnGlyArgThrTyr-Ala- 227  
 QY 722 GCGTGTAAATCACAATTTTAACAT-----TGCAATTATGCACTTTGCTC 675  
 Db 228 -----GlnGlnAspPheArgValGlyGlyThrArgTyrPheAspLeuLeuArgMetP 245  
 QY 674 TGTGCATATCTGCTGTGAGATGTAGCTTTGCTGAGAGGAGTCCAGGTACATGCG 615  
 Db 245 roval-----ArgGlyLeuAspGlyAsp----- 252  
 QY 614 AGTGTCCAGTGCAGATGATCTCTCCACAGGCAAGCCAGCTGTGAGAGTAG-TCG 556  
 Db 253 -----ThrAlaProLeuProProHisThrThrGlnArgGlnThrArgSerA 269  
 QY 555 CTGGTCAGCG-----ACCCTCAGATAGTGTACAGCGCTTCAATATGTGCTGTG 459  
 Db 269 laArgHisProTrpArgGlyLeuArgPheGlyAlaProGlnAlaPheLeuAlaGlyLeuLeu 289  
 QY 498 TGGGATCAGGACCACTTGGCTTCATGCTGAGTCTTGAAGAAAGACAGCTCCGACCTGGG 439  
 Db 289 euAlaAlaValAla-ValGlyThrAlaArgAlaGlyLeuGlnProArgAlaAspMet-- 307  
 QY 438 GGCAGAGTGAAGAGTAACAAGGCTGGATGTTCTCTTGTGAGGCAAGACCATTTCTC 379  
 Db 307 ----- 307  
 QY 378 CTAAATTGACTGCTTGAAGACAGCTATACCTCTCTATACCTCAGACCCCTCTG 319  
 Db 308 -----AlaAlaProPheMet 312  
 QY 318 ACTTCGATGGGGGATCAGTATCTATCAACTCCAGAGAGTCTGCGCATCTGCTTGG 259  
 Db 313 -----Pro-ProGlnPro----- 316  
 QY 258 AGGGCTTTCCTCCAGTGTCTACCTGTGAGAGTGTGCTGCTGTGAGACCAAGGCGCG 199  
 Db 317 -----ProArgAlaHisGlyGlnHisTyrGlnHisTyrGlnHisGlnLeuProH 333  
 QY 198 CACTGGGCAATGGCTTTCACAGATCCACAGCTCAGCCAGGTGAACACTGAAGAGG 139  
 Db 333 eLeuGlyHisAspGlyHisHis-----GlyG1 342  
 QY 138 CAGCTCCGCTGGGTGAGAGTGGCCAGCGGTAGT-----CT 100  
 Db 342 yThrLeuArgValGlyGlnHisHisArgAsnAlaSerAspValLeuProGlyHisTyrPhe 352  
 QY 99 TCCTGGGAGATGTGCTGTCTGGA-----TACACTGTACTCTGACTGTC 54  
 Db 362 uGlnGlyGlyTrpGlyCysTyrAsnLeuSerAspTrpHisGlnGlyThrHisValCys 381

RESULT 14  
 US-08-482-847-1  
 Sequence 1, Application US/08482847  
 Patent No. 5556757



**CORRESPONDENCE ADDRESS:**

ADDRESSEE: Arnold, White & Durkee  
STREET: P.O. Box 4433

STREET: P.O. BOX 4433

CITY: Houston

STATE: Texas

COUNTRY: USA

ZIP: 77210

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk

COMPUTER: IBM PC compatible

OPERATING SYSTEM: PC-DOS/MS-DOS

SOFTWARE: PatentIn Release #1.0, Version #1.30

URGENT APPLICATION DATA:

APPLICATION NUMBER: US/08/690,473

FILING DATE: 26-JUL-1996

CLASSIFICATION: 435

ATTORNEY/AGENT INFORMATION:

NAME: Highlander, Steven L.

REGISTRATION NUMBER: 37.642

REFERENCE/DOCKET NUMBER: ABCD:239

REFERENCE/ DOCUMENT NUMBER: ARCD:233

TELEPHONE: 512/418-3000

TELEPHONE: 312/418-3000  
TELEFAX: 512/474-7577

LEDEKIA: 312/414-1311  
OBMATION FOR SED ID NO:

ORIGIN FOR SEQ ID NO:

SEQUENCE CHARACTERISTICS:

LENGTH: 1298 amino acids

TYPE: amino acid

STRANDEDNESS:

690-473-2

**Alignment Scores:**

Pred. No.:	0.0162	Length:	1298
------------	--------	---------	------

Score:	109.50	Matches:	102
--------	--------	----------	-----

Percent Similarity: 28.81%  
Conservative: 19

Best Local Similarity: 24.29% Mismatches: 131

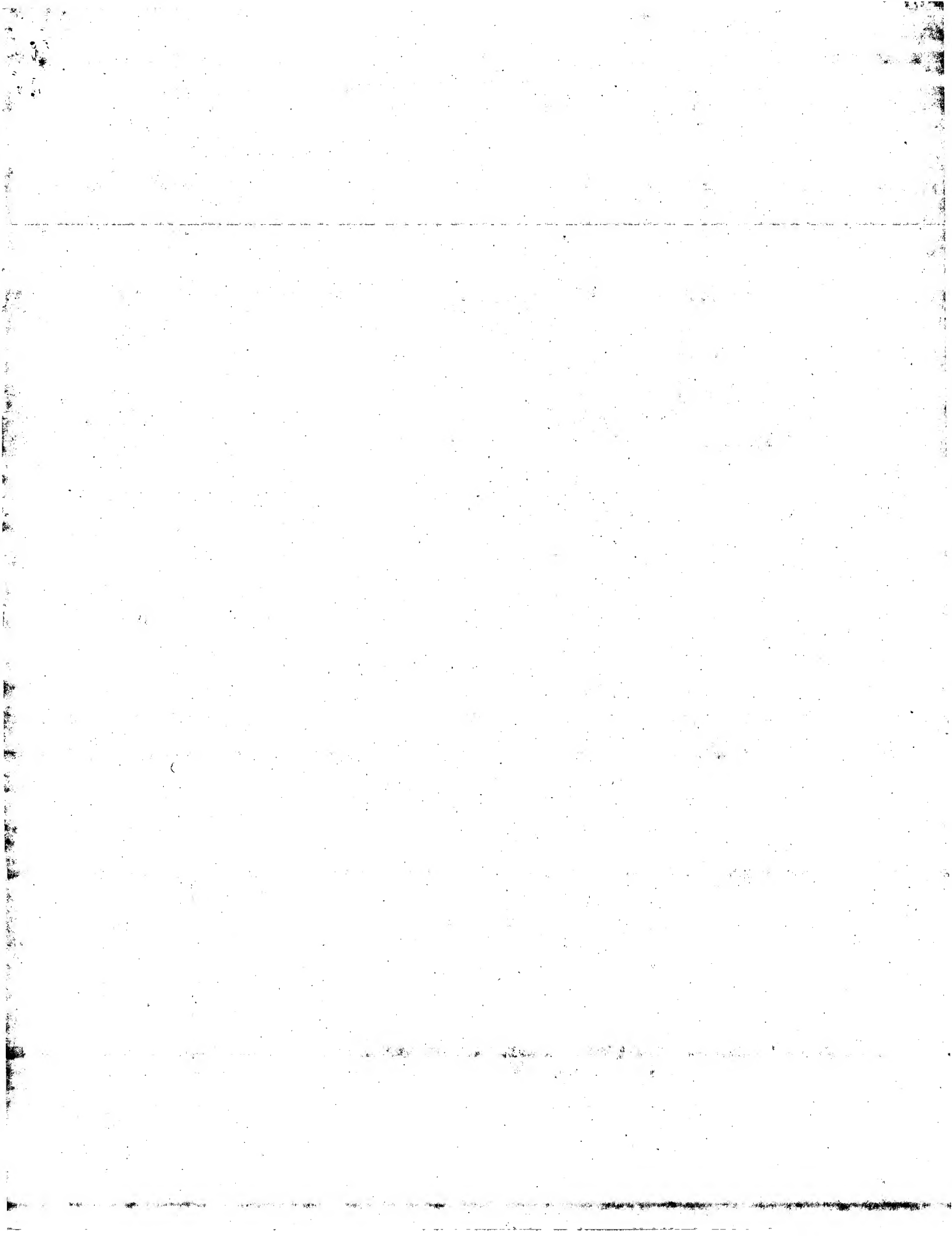
Query Match:	4.28%	Indels:	168
--------------	-------	---------	-----

[illegible]

US-09-989-919-15 (1-1397) x US-08-690-473-2 (1-1298)

[illegible][illegible]

Search completed: June 24, 2003, 14:00:35  
Job time : 50 secs





GenCore version 5.1.6  
Copyright (c) 1993 - 2003 Compugen Ltd.

OM nucleic - protein search, using frame\_plus\_n2p model

Run on: June 24, 2003, 13:52:37 ; Search time 51.5 Seconds  
(without alignments)  
5870.482 Million cell updates/sec

Title: US-09-989-919-15

Perfect score: 2527

Sequence: 1 ggtcgtcgcaccctaccgga.....aaaaaaaaaaagcggtc 1397

Scoring table: BLOSUM62  
Xgapop 10.0 , Xgapext 0.5  
Ygapop 10.0 , Ygapext 0.5  
Delop 6.0 , Delext 7.0

Searched: 417779 seqs, 108206813 residues

Total number of hits satisfying chosen parameters: 835558

Minimum DB seq length: 0  
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Command line parameters:  
-MODEL=frame+ n2p model -DEV=rlp  
-Q=/cgn2\_1/USFTO.spool/US09989919/runat.24062003.102048.19242/app.query.fasta.1.1543  
-DB=Published Applications AA -QFMT=fastcan -SUFFIX=rapb -MINMATCH=0.1  
-LOOPEXT=0 -LOOPEXT=0 -UNITS=bits -START=1 -END=-1 -MATRIX=blosum62  
-TRANS=human40.cdi -LIST=45 -DOCALLIGN=200 -THR SCORE=pct -THR MAX=100  
-THR MIN=0 -ALIGN=1 -MODE=LOCAL -OUTFMT=pcr -NORM=ext -HEADSIZE=500 -MINLEN=0  
-MAXLEN=200000000 -USER=US09989919 @CGN 1.1.17 @runat.24062003.102048.19242  
-NCPU=6 -ICPU=3 -NO MMAP -LARGEQUERY -NRG SCORES=0 -WAIT -DSRBLCK=100  
-LONGLOG -DEV TIMEOUT=120 -WARN TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5  
-FGAPOP=6 -FGAPEXT=7 -YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database : Published Applications AA:

1: /cgn2\_6/prodata/2/pubppa/US08\_NEW\_PUB.pep.\*  
2: /cgn2\_6/prodata/2/pubppa/PCT\_NEW\_PUB.pep.\*  
3: /cgn2\_6/prodata/2/pubppa/US06\_NEW\_PUB.pep.\*  
4: /cgn2\_6/prodata/2/pubppa/US06\_PUBCOMB.pep.\*  
5: /cgn2\_6/prodata/2/pubppa/US07\_NEW\_PUB.pep.\*  
6: /cgn2\_6/prodata/2/pubppa/US07\_PUBCOMB.pep.\*  
7: /cgn2\_6/prodata/2/pubppa/PCTUS\_PUBCOMB.pep.\*  
8: /cgn2\_6/prodata/2/pubppa/US08\_PUBCOMB.pep.\*  
9: /cgn2\_6/prodata/2/pubppa/US09\_NEW\_PUB.pep.\*  
10: /cgn2\_6/prodata/2/pubppa/US09\_PUBCOMB.pep.\*  
11: /cgn2\_6/prodata/2/pubppa/US10\_NEW\_PUB.pep.\*  
12: /cgn2\_6/prodata/2/pubppa/US10\_PUBCOMB.pep.\*  
13: /cgn2\_6/prodata/2/pubppa/US60\_NEW\_PUB.pep.\*  
14: /cgn2\_6/prodata/2/pubppa/US60\_PUBCOMB.pep.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

# SUMMARIES

Result No.	Score	Match	Length	DB	ID	Description
1	939	37.2	175	9	US-09-989-919-84	Sequence 84, Appl
2	283	11.1	52	9	US-09-989-919-83	Sequence 83, Appl
3	134.5	5.3	738	9	US-10-057-487-6	Sequence 6, Appl
4	134.5	5.3	738	10	US-09-978-979-6	Sequence 6, Appl

Result No.	Score	Match	Length	DB	ID	Description
5	134	5.2	4123	9	US-10-213-509-5	Sequence 5, Appl
6	133.5	5.3	1150	9	US-10-123-155-531	Sequence 531, App
7	130.5	5.2	1346	9	US-10-123-155-481	Sequence 481, App
8	122.5	4.8	647	9	US-10-086-464-2	Sequence 2, Appl
9	122.5	4.8	721	9	US-10-086-464-4	Sequence 4, Appl
10	122.5	4.8	721	9	US-10-086-464-5	Sequence 5, Appl
11	121.5	4.7	522	9	US-09-764-868-1138	Sequence 1138, App
12	121.5	4.7	522	9	US-09-764-868-761	Sequence 761, App
13	120.5	4.6	1336	9	US-10-176-847-68	Sequence 68, Appl
14	118	4.6	897	9	US-10-099-899-1	Sequence 1, Appl
15	117	4.6	2447	9	US-10-123-155-291	Sequence 291, App
16	116.5	4.6	2226	9	US-10-123-155-355	Sequence 355, App
17	116	4.5	957	10	US-10-025-380-1065	Sequence 1065, App
18	116	4.5	957	10	US-09-922-217-1065	Sequence 1065, App
19	116	4.5	957	10	US-09-833-263-1065	Sequence 1065, App
20	116	4.5	1259	9	US-10-260-715-8	Sequence 8, Appl
21	115.5	4.6	257	9	US-09-813-153-282	Sequence 282, App
22	114.5	4.5	2782	9	US-10-123-155-205	Sequence 205, App
23	112.5	4.4	376	10	US-09-955-518-20	Sequence 20, Appl
24	112	4.4	1311	9	US-10-103-377C-6	Sequence 6, Appl
25	111.5	4.4	1021	9	US-10-184-644-373	Sequence 373, App
26	111.5	4.4	1021	9	US-10-184-644-373	Sequence 373, App
27	111	4.3	707	9	US-09-919-039-278	Sequence 278, App
28	111	4.4	1231	9	US-10-123-155-279	Sequence 279, App
29	111	4.3	5877	9	US-10-142-515-11	Sequence 11, Appl
30	111	4.3	5935	9	US-10-243-243A-8	Sequence 8, Appl
31	110.5	4.4	1076	9	US-10-184-644-191	Sequence 191, App
32	110.5	4.4	1076	9	US-10-184-644-191	Sequence 191, App
33	110.5	4.4	3266	9	US-10-123-155-211	Sequence 211, App
34	110	4.4	1841	9	US-10-184-644-601	Sequence 601, App
35	110	4.4	1841	9	US-10-123-155-457	Sequence 457, App
36	110	4.4	1841	9	US-10-184-644-601	Sequence 601, App
37	110	4.4	3951	9	US-10-184-644-119	Sequence 119, App
38	110	4.4	3951	9	US-10-184-644-119	Sequence 119, App
39	109.5	4.3	1298	9	US-09-825-288A-2	Sequence 2, Appl
40	109	4.3	1399	9	US-10-184-644-11	Sequence 11, Appl
41	109	4.3	1399	9	US-10-184-644-11	Sequence 11, Appl
42	108.5	4.3	1113	9	US-10-123-155-109	Sequence 109, App
43	108.5	4.2	1203	9	US-10-067-457-3	Sequence 3, Appl
44	108.5	4.3	2185	9	US-10-123-155-437	Sequence 437, App
45	108.5	4.3	2290	9	US-10-123-155-267	Sequence 267, App

## ALIGNMENTS

RESULT 1  
US-09-989-919-84  
Sequence 84, Application US/09989919  
Patent No. US20020164344A1  
GENERAL INFORMATION:  
APPLICANT: Macina, Roberto  
APPLICANT: Recipon, Herve  
APPLICANT: Pluta, Jason  
APPLICANT: Ghosh, Malavika  
APPLICANT: Sun, Yongyang  
APPLICANT: Liu, Chenghua  
TITLE OF INVENTION: Compositions and Methods Relating to Colon Specific Genes and Proteins  
FILE REFERENCE: DEX-0289  
CURRENT APPLICATION NUMBER: US/09/989, 919  
PRIOR FILING DATE: 2001-11-21  
PRIOR APPLICATION NUMBER: 60/252, 505  
NUMBER OF SEQ ID NOS: 124  
SOFTWARE: PatentIn version 3.1  
SEQ ID NO 84  
LENGTH: 175  
TYPE: PRT  
ORGANISM: Homo sapien  
US-09-989-919-84  
Alignment Scores: 2,096-69  
Pred. No.: 939.00  
Length: 175  
Matches: 175

```
Percent Similarity: 99.43% Conservative: 0
Best Local Similarity: 99.43% Mismatches: 0
Query Match: 37.16% Indels: 1
DB: 9 Gaps: 0
```

US-09-989-919-15 (1-1397) X US-09-989-919-84 (1-175)

QY	2	TRGGCTGCACCTGTATACCAGGAGGGGGACGATCTCTGACAAATCCACGGGACAGACAGCTACC	61
Db	1	ValLeuHisLeuIyrArgSerIyGlnIyrLeuGlnHisSerThrIaSerSerSerThr	20
QY	62	GAGTACCACTGTATCCACAGACACACCATCCCCAGAGAAACTACCGCTGCTGCATCC	121
Db	21	GluIyrGlnCysIleProaspSerThrIleProGlnGlnAspIyrArgCysIyrProSer	40
QY	122	TACCAACACAGGGAGCTGCCTCTTACAGTTTCAACCTGGCTAGGCTGTGATGTCGTG	181
Db	41	TyrHisHisGlySerCysLeuLeuSerValPheAsnLeuAlaGlnIValAspValCys	60
QY	182	GAGAGCCATGCCACAGTGTGGGGCTTTGTGGTACCAACACCAACCACTTGACAGGTGAG	241
Db	61	GlnSerHisAlaGlnCysArgIaIaPheValIThrAsnGlnThrIrrThrIrrGlnIyrGlu	80
QY	242	CCAGTGGGAGAGAGCCCTTCCAAAGGAGATGGAGACCTCTCGAGGTTGATAGATAGT	301
Db	81	ProValGlyGlnIaIaLeuProArgIaIaGlnIaGlyProLeuIrrPheGlnIleAspSer	100
QY	302	GATCCCCCATCGGAGTCAAGAGGGGGTGCTGAGGTGATGAGAGAGAGTATACGTGCTTT	361
Db	101	AspProProSerSerIuValaIaArgGlyGlyIaGlnIaIaIaMetArgGlnIaIyrIrrIrrCysLeu	120
QY	362	CAGGAGAGTCAAAATTGGGAGAGATGGCTTGCTCCCAAGAAAGAAACATCCAGCCCTGT	421
Db	121	GlnGlySerGlnIleArgGlnAsnGlyLeuAlaSerArgIaArgAsnIleGlnProCys	140
QY	422	TACCTTCAACCTTGCCCCCAGAGTCGGACGCTGGTCTTTTTTCAAGCTGAATGAGCCA	481
Db	141	TyrLeuSerProIeu-ProProGlyArgGlnLeuValPhePheIyrThrGlyIrrPheSerGly	160
QY	482	ACGGTGCCTGTATCCCAACAAGACCAATATGGAAAGGCGCTTGCC	527
Db	160	IrrValValProAspProAsnIyrHisThrIrrIyrValIyrValaSerSerGly	175

```

RESULT 2
US-09-989-919-83
Sequence 83, Application US/09989919
Patent No. US20020164344A1
GENERAL INFORMATION:
APPLICANT: Macina, Roberto
APPLICANT: Recipon, Hervé
APPLICANT: Pluta, Jason
APPLICANT: Ghosh, Malavika
APPLICANT: Sun, Yongming
APPLICANT: Liu, Chenghua
TITLE OF INVENTION: Compositions and Methods Relating to Colon Specific Genes and Pro
FILE REFERENCE: DEX-0289
CURRENT APPLICATION NUMBER: US/09/989,919
CURRENT FILING DATE: 2001-11-21
PRIOR APPLICATION NUMBER: 60/252,505
PRIOR FILING DATE: 2000-11-22
NUMBER OF SEQ ID NOS: 124
SOFTWARE: PatentIn version 3.1
SEQ ID NO 83
LENGTH: 52
TYPE: PRT
ORGANISM: Homo sapien
US-09-989-919-83

Alignment Scores:
Pred. No.: 2,82e-15
Score: 283.00
Percent Similarity: 100.00%
Best Local Similarity: 100.00%

Length: 52
Matches: 52
Conservative: 0
Mismatches: 0

```

```
Query Match: 11.05% Indels: 0
DB: 9 Gaps: 0
```

US-09-989-919-15 (1-1397) x US-09-989-919-83 (1-52)

Qy	961	ATGAGTATGTCCTCCACGAGAGCCCCACAGCGCTGCTGTGCATGTGCACAGAGACAGACAGG	902
Db	1.	MeAspCysArgPheHisAlaIaIaProThrAlaCysCysGlyMetCysSerSerSerSerArg	20
Qy	901	GGGTTAGTAGTACATTTCACACTATTGTAACACAGAGATGAGGAGTGGCCACAGGAACCGTCC	842
Db	21	GlyPheSerTyrIleLeuThrIleuLeuSerThrValMetGlyLeuProThrGluProSer	40
Qy	841	CAGGAGAGGGGACACAGCCCCAGTGGGAGACACTAGCC	806
Db	41	GlnGlyGlyAlaGlnProProValGlyArgLeuAla	52

```

1 RESULT 3
2 US-10-057-487-6
3 Sequence 6, Application US/10057487
4 Publication No. US20030105313A1
5 GENERAL INFORMATION:
6 APPLICANT: Wyeth
7 TITLE OF INVENTION: Aggricanease Molecules
8 FILE REFERENCE: 08702..0073
9 CURRENT APPLICATION NUMBER: US/10/057,487
10 PRIORITY FILING DATE: 2002-01-25
11 PRIOR APPLICATION NUMBER: 60/241,469
12 PRIOR FILING DATE: 2000-10-18
13 NUMBER OF SEQ ID NOS: 8
14 SOFTWARE: PatentIn version 3.1
15 SEQ ID NO 6
16 LENGTH: 738
17 TYPE: prt
18 ORGANISM: homo sapien
19 FEATURE:
20 NAME/KEY: MISC_FEATURE
21 LOCATION: (43) .. (43)
22 OTHER INFORMATION: unknown amino acid
23 FEATURE:
24 NAME/KEY: MISC_FEATURE
25 LOCATION: (192)..(192)
26 OTHER INFORMATION: unknown amino acid
27 FEATURE:
28 NAME/KEY: MISC_FEATURE
29 LOCATION: (255)..(255)
30 OTHER INFORMATION: unknown amino acid
31 FEATURE:
32 NAME/KEY: MISC_FEATURE
33 LOCATION: (258)..(258)
34 OTHER INFORMATION: unknown amino acid
35 FEATURE:
36 NAME/KEY: MISC_FEATURE
37 LOCATION: (374)..(374)
38 OTHER INFORMATION: unknown amino acid
39 FEATURE:
40 NAME/KEY: MISC_FEATURE
41 LOCATION: (397)..(397)
42 OTHER INFORMATION: unknown amino acid
43 FEATURE:
44 NAME/KEY: MISC_FEATURE
45 LOCATION: (452)..(452)
46 OTHER INFORMATION: unknown amino acid
47 FEATURE:
48 NAME/KEY: MISC_FEATURE
49 LOCATION: (458)..(458)
50 OTHER INFORMATION: unknown amino acid
51 FEATURE:
52 NAME/KEY: MISC_FEATURE
53 LOCATION: (475)..(475)
54 OTHER INFORMATION: unknown amino acid
55 FEATURE:
56 NAME/KEY: MISC_FEATURE

```



Alignment Scores:	
Pred. No.:	0.012
Score:	134.50
Percent Similarity:	32.09%
Best Local Similarity:	24.11%
Query Match:	5.25%
DB:	10
Length:	738
Matches:	46
Conservative:	45
Mismatches:	170
Indels:	214
Gaps:	29

QY	1265	TTCAAACTCCAGAGGAGGGGAAAGACCTGGTTTGGCTGCTTTAGAGCCTTTATTGAAGC	12068
Db	2	PhcelyleuGIunHIsaepGIyAlaProGIIserylCySgIyProserGIyHIsaValMet	21
QY	1205	AAATGAGGCGAAGAAAGGCCCGGGCCCCACCCCGGCCAGGTCT	1158
Db	22	AlaserGIunrIgaIygaIyProIalaPro--AlaserProGIyProIalaIalaIaIaIaIyS	41
QY	1157	CAAACCAACCTGTCTCTCCCTGACCCCACTTTGGCTCAIgcTGAAGTGGC-----A	1107
Db	41	ercys**AlaCysser--AspProserIeuArGIarGIarSerIeuCyetrIyrProIrot	60
QY	1106	CCCTGTGCCCTTGACCCCTGGGCTGGC-----CTGG	1077
Db	60	hIserIalaPro-----AlaGIyAlaIeuValIeuValProIaIaIySerArIeuL	77
QY	1076	GAGTGTCCCTGTGTGGAGGTCAAGATAGCTCCCAAGATACAGATACCCACATCTT	1017
Db	77	euValGIyGIyAlaGIyArG--GIuIeuIeuPheProIeuThrLyGIyHIsaIaserL	96
QY	1016	GGAAGATC-----CCCCGCAAGTCTCTGTCCAGACTAAGCTGAGAAACACTCCAGAT	960
Db	96	ySaIrPheHIsProArGIaIaHIsSer-----	104
QY	959	GGATTGTCCCAAGCAGACGCCCAACCCAGCTGTCTGTG--	925
Db	105	-----SerValProProProProGIyValHIsProGIyThrGIuProGIyIeuS	121
QY	924	-----GGATGTGAAGCAGCAGCAGCAGGAGGCTTAGCTAATTCTCACACTATTGAACAC	870
Db	121	erArGIaIaIeuSerGIaIrArGIeThrLyAlaIeuVal-----	133
QY	869	AGTGAATGGGGCTGCCCAACG-----AACCTCCAGGAGGAGGCAC	828
Db	134	-----TTrpaPProProIaIyProGIaIaProGIySerAlaGIyHIsProArGIaIaHIsL	152

[illegible]





QY 123 ACCACGAGGAGCTGCTCTTCACTGTTCAACCTG---GCTGAGCTGTGATGTC 178  
 |||||  
 Db 458 ThrThrThrGlyCysAlaThr-----CysAlaThrCysAlaThrAlaGlyThrThrGly 475  
 QY 179 TGTGAGAGCATATCCAGTGTGGGCTTTGTGTACCAACAGACCTGAGACAGGT 238  
 |||||  
 Db 476 CysAlaAlaThrThrGlyCysAlaCysThrAlaCysThrAlaThrThrThrThrAla 495  
 QY 239 GAGCCAGTGGAGAGAGCCCTTCCAGAGAGATGGAGAGACCTCTGAGAGTTGATGAT 298  
 |||||  
 Db 496 ThrCysAlaGly----- 499  
 QY 299 AGTATGCCCATCCGAGAGTCAAGAGGGGTGCTGAGGTGATGAGAGAGATACGTGT 358  
 |||||  
 Db 500 -----GlyGlyAla-----ThrCys 504  
 QY 359 CTTCAAGGC---AGTCAATTAAGGAGAGATGTCTTCTCCGAAAGAGAAACATTCAG 415  
 |||||  
 Db 505 ThrGlyGlyCysAlaAlaCysGlyThrAlaGlyAlaAlaGly----- 518  
 QY 416 CCTGTACTCTCACTCTGACCTGCCCCAGGTGGCAGCTGTCTTTTCAAGACTGGATG 475  
 |||||  
 Db 519 -----AlaAlaAlaGlyAla----- 523  
 QY 476 GAGCCAGTGTCTCTGATCCCAACAGACATATGTGAGAGCTGTGACTGAT 535  
 |||||  
 Db 523 ----- 523  
 QY 536 CTGAGGCTGCTGCTGACAGCTGACTATCTCAAGAGCTGGCTGT----- 586  
 |||||  
 Db 524 -----AlaCysAlaAlaAlaGlyAlaAlaCysCysAlaThr 535  
 QY 587 -----GGAGGAGTACT-----TGCACTGGC----- 608  
 |||||  
 Db 536 CysThrGlyAlaAlaGlyThrGlyGlyAla-ThrGlyAlaCysGlyCysThrGlyAlaAl 555  
 QY 609 -----AGCACTGATGTCACTGGAACCC 634  
 |||||  
 Db 555 AGlyAlaThrAlaAlaGlyThrGlyThrGlyAlaAlaAlaCysAlaThrGlyAlaThr 575  
 QY 635 TGCAGACAAAGCTAACCTCCAGACAGAGATGTGACCAAGCA----- 680  
 |||||  
 Db 575 rCysAlaCysAlaAlaThrThrGlyAlaAlaAlaAlaThrGlyCysAlaThrCysCys 595  
 QY 681 -----ACGTGCAATTAATGCCAAATGTAAATGTGAGTTTACAGCTTATGGA 733  
 |||||  
 Db 595 rCysCysThrCysThrGlyAlaThrCys----- 604  
 QY 734 CTGCTGAGCTCTAGTCCAGGAATCATGAGGGGTATGACTGCTCTCCAACTGTGAGGCTG 793  
 |||||  
 Db 605 -----CysCysCysThrGlyAlaCys 612  
 QY 794 TAAGCAAGCTGAGCTAGTCTCCCACTGGGGCTGTGCCCCCTGAGGAGCTTCCT 853  
 |||||  
 Db 612 salathrGlyAlaAlaGlyGlyGlyGlyGlyCysAlaThrAlaThrAlaAlaThr 632  
 QY 854 GGGGAGCCCATCATCTGTTCATTAAGTGTGAGAAATTAAGCCCTGCTC----- 906  
 |||||  
 Db 632 rGly-AlaThrGlyCysCysThr-----ThrCysAlaThrGlyAlaCysAlaGlyAla 649  
 QY 907 -----TGCTGCTGCTGCAAT 922  
 |||||  
 Db 649 laelGlyAlaThrGlyAlaGlyAlaGlyCysThrCysAlaCysCysCysCysThrC 669  
 QY 923 GCCACA-----GCAGGCGGTGGGGCTGTGGGAGCAATTCATCTGAGAGTTC 973  
 |||||  
 Db 669 ys-ThrCysThrGlyAlaAlaGlyGlyCysThrGlyThr----- 682  
 QY 974 TCTCAGCTTAGTCTGAGACAGAGACTTGGCGGGGATGCTCCAGAGATGCGGTGATTC 1033  
 |||||  
 Db 683 -----ThrGlyThrThr 686

QY 1034 GTAACCTGGGAGGCTATCTTGTGACTCCGACAGGGAGACCTCCAGCCAGGAG 1093  
 |||||  
 Db 687 Cys-----ThrGlyCysThrCysCysThrCysAla 697  
 QY 1094 TCAGGGGAGAGTGCACACTTCAGATGAGCCAAACCTGGGTACAGGAGAGTGTG 1153  
 |||||  
 Db 698 AlaGlyAlaAlaAlaThrThrAlaAlaAlaCysAlaThrThrThrGlyThrThrCys 717  
 QY 1154 TTTGAGCCAGAGACTGGGCGGGGTGGG-----CCGGGCTTTCTGCTCATTTGC 1207  
 |||||  
 Db 718 ThrGlyThrGlyThrGlyAlaCysThrGlyCysThrGlyAlaGlyCysAlaThrCys 736  
 RESULT 8  
 US-10-086-464-2  
 ; Sequence 2, Application US/10086464  
 ; Publication No. US20020199218A1  
 GENERAL INFORMATION:  
 ; APPLICANT: GORING, Daphne R. et al.  
 ; TITLE OF INVENTION: PROLINE-RICH EXTENSIN-LIKE RECEPTOR KINASES  
 ; FILE REFERENCE: P 25,762-A USA  
 ; CURRENT APPLICATION NUMBER: US/10/086,464  
 ; PRIORITY FILING DATE: 2002-02-28  
 ; PRIOR APPLICATION NUMBER: US 10/069,304  
 ; PRIOR FILING DATE: 2002-02-19  
 ; PRIOR APPLICATION NUMBER: PCT/CA00/00966  
 ; PRIOR FILING DATE: 2000-08-18  
 ; PRIOR APPLICATION NUMBER: US 60/149,466  
 ; PRIOR FILING DATE: 1999-08-19  
 ; PRIOR APPLICATION NUMBER: US 60/159,122  
 ; PRIOR FILING DATE: 1999-10-13  
 ; NUMBER OF SEQ ID NOS: 27  
 ; SOFTWARE: Patentin Ver. 2.1  
 ; SEQ ID NO 2  
 ; LENGTH: 647  
 ; TYPE: PRT  
 ; ORGANISM: Brassica napus  
 US-10-086-464-2  
 Alignment Scores:  
 Pred. No.: 0.107 Length: 647  
 Score: 122.50 Matches: 91  
 Percent Similarity: 31.28% Conservative: 36  
 Best Local Similarity: 22.41% Mismatches: 129  
 Query Match: 4.79% Indels: 150  
 DB: 9 Gaps: 19  
 US-09-989-919-15 (1-1397) x US-10-086-464-2 (1-647)  
 QY 1188 CCCGCCCCACCCCGCCAGGTCTGCTCAACACACTGCTCCCTGAGCCGAGTC 1129  
 |||||  
 Db 7 ProGlyThrGlySerProProSerProProSerProProSerProProSerProPro 25  
 QY 1128 TTGGCTCATCTGAGGTGTGCAC-----TTGCCCCCTGACCCCTGGGCTGCTGGA 1075  
 |||||  
 Db 26 ProAlaSerAlaProProProProProProProProProProProProProProPro 45  
 QY 1074 GTGTCC-----CCTTGGGAGAGTCAAGATAGCTCCCGCACTGATCAATCAACCA 1021  
 |||||  
 Db 46 ThrSerProProProSerSerAlaGlySerThrProSerAlaProProProProPro 65  
 QY 1020 TCCTGAGCATCC-----CGCAAGTCTCTGTCCAGACTTAAGC 979  
 |||||  
 Db 66 Pro---SerThrProGlySerProProProProProProProProProProProPro 84  
 QY 978 TGAGAGAACTTCACAGATGATGTC-----CCAGCGAGCCCGACCGCTGTG 925  
 |||||  
 Db 85 ThrProGlySerProProAlaProValThrProProThrArgAsnProProProSerVal 104  
 QY 924 GCATGTGCAGCAGACAGCAGAGGGCTTAACTCACTTCTCACACTATTGAACACAGTGA 865  
 |||||  
 Db 105 Pro----- 105  
 QY 864 TGGGCTGCCACGAAACCGTCCAGAGAGGGGACAGCCCGCAGTGGAGAGTACGCT 805







QY 444 CTGGGGGAGAGAGTGAAGTAAAGGGCTGATGTTCTCTTCTGAGGAGACCA 385  
 Db 195 -----TriprarglnInlha 199  
 QY 384 TTCTCCCTAATTGACTGCTTGAAGACAGTATACCTCTCTCATCACTCAGACACC 325  
 Db 199 naLa ----- 200  
 QY 324 CCTCTACTTCGATGGGAGATCACTATCATCACTCCAGAGAGTCCGCTCATCTCC 265  
 Db 201 -----ThnProProSerAspHisValValInrSer-----LeuProProPr 214  
 QY 264 CTGGAAGGGCTTCTCCCACTGCTCAGCTGTCAGAGTGTCTGTGTGAGACACAAG 205  
 Db 214 oProLySalProSerPro-----Proargl 223  
 QY 204 GCGCCGACACTGGGAGCTCTTCAACAGATCCAGCTCAGCCAGTTGAACACTGAA 145  
 Db 223 nProProPro-----ProProProProPhemeSerSerSe 236  
 QY 144 AGGAGCAGCTCCCGTGTGTAGATGAGCCAGAGGATGCTTCTCTGGGGAGTGTG 85  
 Db 236 rGlyGlySerAspTyrSerAspArgProValLeuProProProSerProGly-----Le 254  
 QY 84 CTGTCTGGATACACTGTAAGTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 25  
 Db 254 uValLeuGlyPheSerLeuSerThrPhe-----ThnTyrGlnGlnLeuVal 269  
 QY 24 CCGCTCCGCTACAGT 9  
 Db 269 aArgAlaThrAsnGly 274

RESULT 10

US-10-086-464-5  
 : Sequence 5, Application US/10086464  
 : Publication No. US20020199218A1  
 : GENERAL INFORMATION:  
 : APPLICANT: GORING, Daphne R. et al.  
 : TITLE OF INVENTION: PROLINE-RICH EXTENSIN-LIKE RECEPTOR KINASES  
 : FILE REFERENCE: P 25,763-A USA  
 : CURRENT APPLICATION NUMBER: US/10/086,464  
 : PRIOR FILING DATE: 2002-02-28  
 : PRIOR APPLICATION NUMBER: US 10/069,304  
 : PRIOR FILING DATE: 2002-02-19  
 : PRIOR APPLICATION NUMBER: PCT/CA00/00966  
 : PRIOR FILING DATE: 2000-08-18  
 : PRIOR APPLICATION NUMBER: US 60/149,466  
 : PRIOR FILING DATE: 1999-08-19  
 : PRIOR APPLICATION NUMBER: US 60/159,122  
 : PRIOR FILING DATE: 1999-10-13  
 : NUMBER OF SEQ ID NOS: 27  
 : SOFTWARE: PatentIn Ver. 2.1  
 : SEQ ID NO 5  
 : LENGTH: 721  
 : TYPE: PR1  
 : ORGANISM: Brassica napus  
 : US-10-086-464-5

Alignment Scores:  
 Pred. No.: 0.11 Length: 721  
 Score: 122.50 Matches: 91  
 Percent Similarity: 31.28% Conservative: 36  
 Best Local Similarity: 22.41% Mismatches: 129  
 Query Match: 4.79% Indels: 150  
 Db: 9 Gaps: 19

US-09-989-919-15 (1-1397) x US-10-086-464-5 (1-721)

QY 1188 CCGGGCCCAACCCCGCCAGGCTCTGCTCAACACACACTGCTCCGACCCAGT 1129  
 Db 37 ProGlyThnGlySerProProSerProProSerAsnSerThr-----ThnThrThnProPro 55

QY 1128 TTGGCTCATGCTAGAGTGTGACCC-----TGTGCCCTGACCCCTGGGCTGGCTGGGA 1075  
 Db 56 ProAlaSerAlaProProProThnThrProSerSerProProProSerThrIlePro 75  
 QY 1074 GTGTCC-----CCTGTGGGAGGTGAGATAGCTCCCGAGTACAGATATACCCCA 1021  
 Db 76 ThnSerProProSerSerSerAspSerThrProSerAlaProProProSerProProThn 95  
 QY 1020 TCTTGAGCATCCCC-----CGCCAGTCTCTGTCAGACCTAAGC 979  
 Db 96 Pro-----SerThnProGlySerProProProLeuProInProSerProProAlaProThn 114  
 QY 978 TGAGAACACTCCACAGATGATGTC-----CCGACGAGCCCGCCAGCTGCTGTG 925  
 Db 115 ThnProGlySerProProAlaProValThnProProThnArgAsnProProSerVal 134  
 QY 924 GCATGTGACGACGACGACGAGGGGCTTAACTACATTCTACACATTGAACACAGTGA 865  
 Db 135 Pro----- 135  
 QY 864 TGGGGCTGCCACGACCGCTCCAGAGGGGACAGCCCGCAGTGGGAGACTAGCT 805  
 Db 136 ---GlyProProSerAsnProSerArgGlnGlySerProArg-----Pro 150  
 QY 804 GAGCTTCTTACAGCCGACAGGAGTTGAGAGGACAGTCAACCCCATGATTCCTGAGCTA 745  
 Db 151 ProSerSerProSerProPro-----SerProSerSerAspGlyLeu 164  
 QY 744 GAGCCAGCAGTCCCATAGCTAGCTGTAACCTCACTATTGAATTTGCAATTATTCG 685  
 Db 165 SerThnGlyValValValGlyIleAlaIleGlyIly-----ValAla 178  
 QY 684 ACCTTGTCTGCTGATCATCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 625  
 Db 179 LeuLeuValIleValIleThrLeuIleCysLeu-----LeuCysIleValIleValArg 195  
 QY 624 GGTACATCAGTGTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 565  
 Db 196 ArgAspGlnGlnAspAlaIleTyrValProProProProProProGlyProLySalA 214  
 QY 564 GATATGCTAGCTGTGACGCGGAGCTCAGATAGTCAAGCCAGAGGCTTCACTATGTG 505  
 Db 214 ----- 214  
 QY 504 TCTTGTGGGATCAGGAGACACTGTGCTCATCCAGCTTTGAAGAACACAGCTGCCGAC 445  
 Db 215 -----GlyGlyProTyrGlyGlnGln-Gln----- 224  
 QY 444 CTGGGGGAGAGTGAAGTAAACAGGCTGATGTTCTTCTGAGGAGACACCA 385  
 Db 225 -----TriprarglnInlha 229  
 QY 384 TTCTCCCTAATTGACTGCTTGAAGACGTAATCTCTCTCATCACTCAGACACC 325  
 Db 229 naLa ----- 230  
 QY 324 CCTCTACTTCGATGGGAGATCACTATCATCACTCCAGAGAGTCTGCTCATCTCC 265  
 Db 231 -----ThnProProSerAspHisValValInrSer-----LeuProProPr 244  
 QY 264 CTGGAAGGGCTTCTCCCACTGCTCAGCTGTCAGAGTGTCTGTGTGAGACACAAG 205  
 Db 244 oProLySalProSerPro-----Proargl 253  
 QY 204 GCGCCGACACTGGGAGTGTCTTCAACAGATCCAGCTCAGCCAGTTGAACACTGAA 145  
 Db 253 nProProPro-----ProProProProPhemeSerSerSe 266  
 QY 144 AGGAGCAGCTCCCGTGTGTAGATGAGCCAGAGGATGCTTCTCTGGGGAGTGTG 85  
 Db 266 rGlyGlySerAspTyrSerAspArgProValLeuProProProSerProGly-----Le 284  
 QY 84 CTGTCTGGATACACTGTAAGTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 25



## Alignment Scores:

Pred. No.: 0.122 Length: 524  
 Score: 121.50 Matches: 105  
 Percent Similarity: 32.63% Conservative: 35  
 Best Local Similarity: 24.48% Mismatches: 128  
 Query Match: 4.75% Indels: 161  
 DB: 9 Gaps: 27

US-09-989-919-15 (1-1397) x US-09-764-868-761 (1-524)

QY 1191 GGGCCCCGGCCCCACCCCG---GCCCGAGTCTGGCTCAACACACACCTGCTCCGAC 1135  
 DB 78 GlySerGlyProGluProLeuAlaProSerPro-----ValSer 90  
 QY 1134 CCAAGTCTGGCTCATGCTGAGGTGTGACCTGCCCCGTGACCCCTGGGCTGGCGGA 1075  
 DB 91 ProThrPhePro-----ProSerSerProSerAspTrpPro---GlnGlu 104  
 QY 1074 GTGTCCCTGTCCGAGAGTCAAGATAGCTCCCGAGTACAGAAATCACCCACATCC--- 1018  
 DB 105 ArgSerProGlyGlnHisSerAspGlyAlaSerProArgSerProValProThrTrpLeu 124  
 QY 1017 -----TGAGCATCCCCCGCCCAAGTCTCTGTCAGACCTAAGC 979  
 DB 125 ProGlyLeuArgHisAlaProTrpGlnGlyProArgGly-----ProProAspSer 141  
 QY 978 TGAGAGAACACTCCAGATGATGATTTGCCACAGCAG-----CCC 940  
 DB 142 ProAspGlySerProLeuThrProValProSerGlnMetProTrpLeuValAlaSerPro 161  
 QY 939 CCAACCGCTGCTGTGCATGTGACAGCAGCAGAGGGCTTACATTCATCTCACAC 880  
 DB 162 GluProPro-----GlnSerSerProThrProAlaPheProLeuAlaAlaSer 177  
 QY 879 TATTGAACACAGATGATGGGCTGCCA-----CGAACCTCCACAGGAG 835  
 DB 178 TyrAspThrAsnGlyLeuSerGlnProProLeuProGlyLysArgHisLeuProGlyPro 197  
 QY 834 GGGACACAGCCCC---CAGTGGGAGAGACTAGCTGAGCTTCTTACAGCCACAGGGTGG 778  
 DB 198 GlyGlnGlnProGlyProTrpGly-----ProGlnGlnAlaSerSerProAla----- 213  
 QY 777 AGAGGACGATACCCCATGATCTCTGACTAGAGCAGCAGAGTCCCATAGCTAGAGCTG 718  
 DB 214 ArgGly----- 215  
 QY 717 GTAACACTCACATTTTAAACATTTGGCATTTATGACAGTTTGCTGTCATCTGTCTGT 658  
 DB 216 -----IleSerHisIleValThrPheAlaProLeuLeuSer----- 227  
 QY 657 CTGGAGTGTAGCTTTGTCTGACAGGGGTCCACAGGTACATGACGTCTGCCAGTGCAG 598  
 DB 228 -----Asp 228  
 QY 597 TCATCCCTCCACAGCAGCAGCCAGCTGTGAGATAGTCACTGTCGACCGACCTC 538  
 DB 229 AsnValProGlnThrProGluProProThrGlnIleSerGlnSerAsnValLysPheVal 248  
 QY 537 AGATAGGTCAAGCAGGCTTCAATATGTGTCTTTGGTGGATCAGGAGCACTTGGC 478  
 DB 249 GlnAspThrSerLys-----PheTrp-TyrlsPheHisLeu 261  
 QY 477 TCCATCC-----AGTCTGAAAAGAACAGCTGCCACCTGGGGGGCAGAGGTGAG 427  
 DB 261 ValGAspGlnAlaIleAlaLeuLeuLysAspLysAspPro----- 274  
 QY 426 AGGTAACAGGGGTGATGTTCTTCTCTGAGGACAGACCATCTCCCTAATTTGACTG 367  
 DB 275 -----GlyValaPheLeuIle-----ArgAspSerHisSer----- 284  
 QY 366 CCTTGAAGACAGCATATACCTCTCTCATCATCCTCAGACACCCCTCTGACTTCCGATGG 307  
 DB 366 ----- 307

DB 285 -PheGlnGlyAlaTyr-GlyLeuAlaLeuLysValAlaThrPro----- 298  
 QY 306 GGATCACTATCTATCAACCTCCAGAGAGGTCTGCATCTCCCTGGAAGGC----- 254  
 DB 299 -----ProProSerAlaGln-----ProTrpLysGlyAspPro 310  
 QY 253 -----TTCTCCACCTGGCTC-----ACCTGTCCAGGTGT-----CTGATG 217  
 DB 310 alGlnGlnLeuValArgHisPheLeuIleGluThrGlyProLysGlyValLysIleLysG 330  
 QY 216 GTGACCAACAAAGCCCGCAGACCTGGC-----ATGCTCTCACACATCCACAGCTCA 163  
 DB 330 LysProSerGlnProLysPheGlySerLeuSerAlaLeuValSerGlnHisSerLys 350  
 QY 162 GCCAGGT----- 145  
 DB 350 etProIleSerLeuProCysCysLeuArgIleProSerLysAspProLeuGlnGluThrP 370  
 QY 144 AGAGGACACTCCCTCGTGTGATAGATGCGCACAGCAGGATGTTCTTGGGGATGGT 85  
 DB 370 roGlu---AlaProValProThrAsnMetSerThrAlaAlaPheLeuArgGlnGlyA 389  
 QY 84 CTGTCTGGATACATCTGTACTC 62  
 DB 389 laAlaCysSerValLeuTyrLeu 396  
 RESULT 13  
 US-10-176-847-68  
 ; Sequence 68, Application US/10176847  
 ; Publication No. US20030068636A1  
 ; GENERAL INFORMATION:  
 ; APPLICANT: Velby, Pectet Ole  
 ; TITLE OF INVENTION: COMPOSITIONS, KITS, AND METHODS FOR  
 ; IDENTIFICATION, ASSESSMENT, PREVENTION, AND THERAPY OF BREAST  
 ; FILE REFERENCE: MRI-039  
 ; CURRENT APPLICATION NUMBER: US/10/176,847  
 ; CURRENT FILING DATE: 2002-06-21  
 ; NUMBER OF SEQ ID NOS: 112  
 ; SOFTWARE: FastSeq for Windows Version 4.0  
 ; SEQ ID NO 68  
 ; LENGTH: 1236  
 ; TYPE: PRT  
 ; ORGANISM: Homo sapiens  
 US-10-176-847-68  
 Alignment Scores:  
 Pred. No.: 0.186 Length: 1236  
 Score: 120.50 Matches: 88  
 Percent Similarity: 32.37% Conservative: 35  
 Best Local Similarity: 23.16% Mismatches: 135  
 Query Match: 4.71% Indels: 122  
 DB: 9 Gaps: 19  
 US-09-989-919-15 (1-1397) x US-10-176-847-68 (1-1236)  
 QY 1199 GGCAGAAAGCCCGCGCCACCCCG-----GCCCGAGTCTG 1161  
 DB 926 GlyArgAlaAlaProProSerProCysProThrProArgSerGlyProSerProCys 945  
 QY 1160 GCTCAAC-----ACACTGCTCCCTGACCCCACTTGTG----- 1125  
 DB 946 LeuProThrProAspProProGluProSerProThrGlyTrpLysProProAspGly 965  
 QY 1124 -----CTATGCTGAGGTGTGCACT-----CTGCCCTGACCTCTGG 1086  
 DB 966 GlyArgAlaAlaLeuValArgAlaProGlnProGlyArgProProThrProGly 985  
 QY 1085 CTGGCTGGAGATGCTCCCTGTCCGAGGTCAAGATACCTCCAGGTACAGATC 1026  
 DB 986 -----ProProLeuSerAspValSerArg-----ValSerArgArg 997  
 QY 1025 CCACATCTGAGCATCCCGCCCAAGTCTCTGTCAGACC---TAACGTGAGAGACA 969

```

Db      998 Pro-----AlatrpGluAlaArgTrpProValArgThrIleHisCysGlyArgHis 1014
QY      968 CTCACAGATGATGTCCTCCACAGCCAGCCCGCTGCTGGCATGTGCAGCAGCAG 909
Db      1015 LeuSerAlaSerIleArgProLeuSerProAlaArg-----1026
QY      908 CACAGAGGGCTTTAGCTACTTCTCACATATTAACACATGAGGGCTGCCACCGA 849
Db      1027 -----CysHisIlyrSerSerPheProArgAlaAspArgSerGlyArgPro 1041
QY      848 ACCGTCACAGAGAGGGGACACAGCCCGCTGAGGAGACTGCTGAGCTTGTATACAGC 789
Db      1042 PheLeuProLeuPheProGluProProGluLeuGln---AspLeuProLeuLeuGlyTr 1060
QY      788 CACAGAGGTG---GAGAGGAGCATATACCCCATGATCTCTGAGATAGAGCCAGCATC 732
Db      1060 ogIuGlnLeuAlaArgArgGlnAlaLeuHisAlaAlaIleProAlaArgGlySerArg-- 1079
QY      731 CCATAGCTAGGCTGTGTAACCTCACATTTTAACATTTTGACATTTGCTCTG 672
Db      1079 -----1079
QY      671 TCACATCTGTCTGTGGAGATTTAGCTTTGTCTGCAGGGGTTCCAGTGCATGACATG 612
Db      1080 -----ProArg---HisAlaIse 1084
QY      611 GCTGCCAGTGCAGTACCTCCCTCCACAGGACCCAGCTGCTGAGATGATGACGTG 552
Db      1084 rIeuProSerSerValAlaGlnAlaPheAlaArgProSer-----SerIeuProAlaG 1102
QY      551 TCAGCCGAGCCCTCAGATAGTACGCCAGAGGCTTACATATGTGCTTTGGGATC 492
Db      1102 YCysThrGlyProIleCysAlaArgProAspGlyHisSerAlaCysArgArgLeuAlaG 1122
QY      491 AGGACACACATGGCTTCATCCAGCTTGAAAGAACACAGAGCTGCGAGGAGGAG 432
Db      1122 nAlaGlnSerMetCysLeuProIleIleArgGln---AlaCysGlnGlnGlyGln-- 1140
QY      431 GTGAGAGTAAACAGGGGTGATGTTCTCTTTCGAGGACAGACATTTCCCTAATT 372
Db      1141 -----AlaGlyAl 1143
QY      371 GACTGCTTGAAGACAGTATACCTCTCTCATCACTCAGACCCCTCTGACTCCG 312
Db      1143 aProIleAlaTrpGlnHisArgGlnHisValCysLeuHisAlaHisAlaHisIleuProPheCy 1163
QY      311 ATGGGGATGATCATATATCACTCCAGAGAGGTCCTGCATCTCCCTTGAAGGACT 252
Db      1163 sTrpGlyAlaValCysProHisIleuProProCysAlaSerHis-----1177
QY      251 CTCCTACTGGCTCCTGTCCAGGTGTCTGGTTGTGACACCAAGGCCGACACTGG 192
Db      1178 -GlySerTrpLeuSerGlyAlaTrpGlyProLeuGly---HisArgIleArgThrLeuG 1196
QY      191 CATG-----GCTCTACAGACATTCACAGCTCAGCCAG 158
Db      1196 yLeuGlyThrGlyTrpArgAspSerGlyGlyLeuAspGlnIleSerSerValAlaArg 1215

```

```

; PRIOR FILING DATE: 2000-09-15
; NUMBER OF SEQ ID NOS: 47
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 1
; LENGTH: 897
; TYPE: PR
; ORGANISM: Homo sapiens
US-10-099-895-1

Alignment Scores:
Pred. No.: 0.275 Length: 897
Score: 118.00 Matches: 104
Percent Similarity: 34.37% Conservative: 51
Best Local Similarity: 23.06% Mismatches: 152
Query Match: 4.61% Indels: 144
Gaps: 22

US-09-989-919-15 (1-1397) x US-10-099-895-1 (1-897)

QY      1221 GGCTTTCATGTAAGCAATATGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGTCT 1162
Db      464 GlyIleTrpGlyTrpArgLeuArgArgIleTrpGlnGlnIleProAsnProSerIys 483
QY      1161 GGC-----TCAACACACACTGCTCTCCAGCCAGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1111
Db      484 SerHisIleuPheGlnAsnIleSerAlaGlnLeuTrpProProGlySerMetSerAlaPhe 503
QY      1110 TGCACCTGTCGCT-----GACCTTGAGCTGAGCTG-----GGA 1075
Db      504 ThrSerGlySerProProHisIleGlnGlyProTrpGlySerArgPheProGlnLeuGlnGly 523
QY      1074 GTGTCCCTGTC-----GGAGGTCAGAGATAGCTCCCGAGGATACAGATACCCACA 1021
Db      524 ValPheProValGlyPheGlyAspSerGlnVal-----SerProLeu 537
QY      1020 TCGTGAGAGATCCCGGACCAAGTCTCTGCTCCAGACTTAAGCTGAGAGAACACTCCACGA 961
Db      538 ThrIleGlnAspProIleHisValCysAspProProSerGlyProAspThrProAla 557
QY      960 TGGATT--GTCCCAAG-----CACCCCAACCGCTGCTGTGGCA 922
Db      558 AlaSerAspLeuProThrGlnGlnProProSerProGlnProGlnProAlaAlaSer 577
QY      921 TGTGACAGCAGCAGCAGCAGGAGGCTTAGCTA-----891
Db      578 HisThr-ProGlnIleValGlnAlaSerSerPheAspPheAsnGlyProTrpLeuGlyProPr 597
QY      890 -CATTTCACTATTAAGACAGAGTATGGGCTGCCACGAGAACCGTCCAGAGAGGG 832
Db      597 HisSerArgSer-LeuProAspIleLeuGlnGlnProGlnProProGlnGlnGlyS 617
QY      831 CACAG-----CCCGCAGTGGAGACTGAGCTGAGCTTGTATACAGCCCAAGGTTGG 778
Db      617 erGlnIleSerProProProGlySerIleGlnIleuGlyLeuGlySerLeu---ProAlaGlyG 636
QY      777 AGAGGAGTCATACCCCATGATTCCTGACTAGAGAGCAGAGAGTCCATAGCTAGGCTG 718
Db      636 InValGlnLeuValPro-----641
QY      717 GTAACCTACATTTTAACATTTTGACATTTGACAGTTTGTCTGTGTCACATCTGTCTGT 658
Db      641 -----641
QY      657 CTGGAGATGATGTTGTCTGCAGGGGTTCCAGGTGATGATGATGATGATGATGATGATGATG 598
Db      642 -----LeuAlaGlnAlaMetGlyProGlyGln-----AlaValGlnValGln 656
QY      597 TCATCTCCCTCCACAGCAGCCAGCTGCTGAGAGATGACAGTGTGCTAGCGAGGCTTC 538
Db      656 rArgProSerGlnGly-----AlaAlaGlySerProSerLeuG 669
QY      537 AGATAGTACAGCAGAGGCTTCAATATGATGTGCTTTGGGATCAGGAGCACCATTGGC 478

```

# RESULT 14

US-10-099-895-1  
 Sequence 1, Application US/10099895  
 Patent No. US20020177166A1  
 GENERAL INFORMATION:  
 APPLICANT: BERNDT, Michael C  
 APPLICANT: STOMSKI, Frank C  
 APPLICANT: LOPEZ, Angel F  
 APPLICANT: GUTHRIE, Mark A  
 TITLE OF INVENTION: A BINDING MOTIF OF A RECEPTOR  
 FILE REFERENCE: 3991/0K379  
 CURRENT APPLICATION NUMBER: US/10/099,895  
 CURRENT FILING DATE: 2002-03-14  
 PRIOR APPLICATION NUMBER: PCT/AU00/01118

```

Db      669  IuSerGlyIy-----GlyProAlaProProAlaL 679
QY      477  TCCATCCAGTCTTGAAMAAAGACAGCTGCCAGCTGGGGGAGAGGTGAGGTAAACAG 418
Db      679  euGlyPro-----ArgValGlyGlyInaSerGlnLysAspSer 692
QY      417  GGCTGGATGTTCTCTTCTTCTGAGGCAAGCAATTCCTCCCTAAATTGATGCTTGAAGA 358
Db      692  roValAlaIleProMetSerSerGlyAspThrGlnaAspProGlyValAlaSerGlyTyrV 712
QY      357  CAGGTATACCTCTCTCCATCACCCTGACACCCCTCTGACTTCCCATGGGGATCACTA 298
Db      712  alSer-SerAlaAspLeuVal-----PheThrProaenSerGlyAlaSer 726
QY      297  TCTATCAACCTC-----CAGAGAGTCTCTGCTCATCTCCCTT 262
Db      727  SerValSerLeuValProSerLeuGlyLeuProSerAspGlnThrProSerLeuCySpro 746
QY      261  GGAAGGGCTTCT---CCCACTGGCTCACTGTCAGAGT----- 226
Db      747  GlyLeuAlaSerGlyProGlyAlaProGlyProValLysSerGlyPheGlnGlyTyr 766
QY      225  GTCTGTTGTGTGACCAAAAGGCCGACACTGGGCGATGCTCTCACAGATCCACAGCC 166
Db      767  ValGlnLeuProProIleGlnGlyArg-SerProArgSerProArgAsnAsnProValPr 786
QY      165  T-----CAGCAGGTTGAACACTGAAGAGGAGGAGCTCCG 130
Db      786  oProGlnAlaLysSerProValLeuAsnProGly-----GluAr-799
QY      129  TGGTGTAGAGATGGCCAGCAGCGGTACTCTTCTGGGGGATGTGCTGTCTGGAGTACAC 70
Db      799  gProAlaAspValSerProThrSerProGlnProGlnGlyLeuLeuValLeuGlnGlnVa 819
QY      69  TGCTACTCGGTACTGCTGCTTCCG 45
Db      819  IGIAsp---TyrCySpheLeuPro 826

RESULT 15
US-10-123-155-291
; Sequence 291, Application US/10123155
; Publication No. US20030068794A1
; GENERAL INFORMATION:
; APPLICANT: Baker, Kevin P.
; APPLICANT: Beresini, Maureen
; APPLICANT: DeForge, Laura
; APPLICANT: Desnoyers, Luc
; APPLICANT: Filvaroff, Ellen
; APPLICANT: Gao, Wei-Qiang
; APPLICANT: Gerlitsen, Mary E.
; APPLICANT: Goddard, Audrey
; APPLICANT: Godowski, Paul J.
; APPLICANT: Gurney, Austin L.
; APPLICANT: Sherwood, Steven
; APPLICANT: Smith, Victoria
; APPLICANT: Stewart, Timothy A.
; APPLICANT: Tuma, Daniel
; APPLICANT: Wood, William
; APPLICANT: Watanabe, Colin K
; APPLICANT: Zhang, Zemin
; TITLE OF INVENTION: SECRETED AND TRANSMEMBRANE POLYPEPTIDES AND NUCLEIC
; FILE REFERENCE: P3330R1G30
; CURRENT APPLICATION NUMBER: US/10/123,155
; PRIORITY FILING DATE: 2002-04-15
; Prior Application removed - See Palm or File Wrapper
; NUMBER OF SEQ ID NOS: 550
; SEQ ID NO 291
; LENGTH: 2447
; TYPE: DNA
; ORGANISM: Homo Sapien
US-10-123-155-291

```

```

Alignment Scores:
Pred. No.: 0.437 Length: 2447
Score: 117.00 Matches: 63
Percent Similarity: 32.13% Conservative: 8
Best Local Similarity: 28.51% Mismatches: 66
Query Match: 4.63% Indels: 84
DB: 9 Gaps: 12

US-09-989-919-15 (1-1397) x US-10-123-155-291 (1-2447)

QY      591  GGAGTACTTGACCTGGACGACT-----GCATGTCACTGGGAACCTCTGA 638
Db      1841  GlyAlaCySthrGlyThrGlyCysAlaAlaAlaCysThrAlaGlyThrGlyAla 1860
QY      639  GACAAAGCTAATCCACAGACAGATGTGACAGCAACG----- 683
Db      1861  CySthrAlaGly-----GlyAlaThrAlaGlyAlaAlaThr 1872
QY      684  -----TGCATTAATGCCMAATGTTAAATGTGAGTTTACAGGC 722
Db      1873  ThrThrGlyAlaThrCySthrCysCysCysCysThrAlaAlaCys----- 1885
QY      723  TACCTATGGAGCTGCTGCTCTCTAGTCCAGAAATCATGGGGGTATACCTCTCTCAAC 782
Db      1886  -----ThrCys-----AlaCysThrGlyThr 1892
QY      783  CCTGTGGGCTGTAAAGCAAGCTAGCTAGCTCCCACTGGGGGCTGTGCCCTCCCTG 842
Db      1893  CySthrGlyCysGlyGlyThrGlyCysThrCysAlaAlaThrThrGlyCysThr 1909
QY      843  GACGGTTCGTTGGGAGCCCATCACTGTTCATATAGTGAATGTAGCTAAAGCC 902
Db      1910  ---GlyCysThr-AlaAlaCysAlaGlyCysAlaThrThrGlyCysCysThrGlyThrG 1928
QY      903  CTGC---TGTCTGTCTGCATGCTCACA---GCAGCGGTGGGGCTGCTGGGACAA 956
Db      1928  yCysThrCysThrCysCysThrCys-ThrCysAlaGlyGlyGlyCysAlaGlyCysAl 1948
QY      957  TCCATGCTGAGAGTCTCTCAAGCTTAGCTGAGACAGAGACTTGGGGGGATGCTCC 1016
Db      1948  1a---ThrGly-CysThrAlaAlaCysGly-----GlyGlyGlyCysGly 1961
QY      1017  AGGATGTGGGTATTTCTGTACTGTGGGAGGCTATCTGTGACCTCCGACAGGGACATC 1076
Db      1962  -AlaCysGlyThrCysCysThr----- 1968
QY      1077  CCAGGCCACCCAGGGGCTCAGGGGCGAGAGTGCACACTTCAGCATGACCAAGACTGGGG 1136
Db      1969  -----AlaAlaThrCysCysAlaAlaCysThrG 1978
QY      1137  TCAGGAGCAGAGTGTGTTGAGCCAGACCTGGGGCGGG---GTTGGGGCGCGG 1168
Db      1978  yGlyGlyAlaGlyAlaAlaGlyCysCysThrCysAlaGlyThrGlyGlyThrGly 1996

```

Search completed: June 24, 2003, 14:02:23  
Job time : 77.5 secs





OY	1141	CCTGACCCCAAGCTCTTGCTCAATGCTAGAGTGTGACACTCTGTGCCCT	1098
Db	1805	ProtaGpProGln- ProProProProGlnIsetysCysAlaLeuProProAgSerLeuPr	1824
OY	1095	----GACCCCTGGGGCTGGAGGTGTCCCTCTCGGAGAGTGCAGATAGACCTCCCC	1040
Db	1824	oSerAspProPheSerArgValProValSerProIlnberGlnSerSerGlnSerPr	1844
OY	1039	AGGTACAGAAATACCCACATCTCTGAGACATCCCCGCCAAGTCTCTGTCAGACTAAG	980
Db	1844	oleuThrProArgProLeuSer-----AlaGluAlaPheCysProSer-----	1858
OY	979	CTGAGAGAAACCTCCACAGATGATGTCTCCACAGGAGGCCACCGCTGCTGTGCATG	920
Db	1859	----ProValThrProArgPhe-----GlnSerProAspPro-----	1865
OY	919	TGCAGACAGCAGCAGAGGGGCTTTAGTACATTCACACTATTGAACACAGTATGGG	860
Db	1870	-----TySe	1871
OY	859	CTGCCACGGAACCGTCCACAGGAGGG-----CACAGCCCCCACTG	818
Db	1871	rArgProProSerArgProGlnSerArgAspProPheAlaProIlnbSlusProProAr	1891
OY	817	GGGAGACTAGCTGAGCTTGTCTTACAGCCACAGGGGTGGAGAGGCACT-----CATAC	764
Db	1891	gProGlnProProGluValAlaPheLys-----AlaGlySerLeuAlaIstH	1907
OY	763	CCCCATG---ATTCCGTGACTAGAGGACGACAGCTCCCATAGCTAG--GCTGTAACCTAC	708
Db	1907	rSerLeuAlaGlyGlyPheProAlaAlaLeuProAlaGlyProAlaGlyGluLeu-	1926
OY	707	ATTTAACTTTGGATATTGACAGCTTTGCTGCTGACATCTCTGTCTGGAGATTG	648
Db	1927	-----HisAlaLysValProSerGlyGlnProProAsnPhaValaLagSerProGlyTh	1944
OY	647	AGCTTTGCTCGAGGGGTTCCCAAGTACATGCAGTGTCTGCCAGTGCACATCACT---	593
Db	1944	rGlyAlaPheValGlyThrProSer-----PrometaArgPheThrPhePr	1959
OY	592	-----CCCTCACAGGACGAGCCAGCTG	570
Db	1959	oGlnAlaValGlyLupProSerLeuLysProProValProGlnProGlyLeuProProPr	1979
OY	569	CTGAGATAGTCAAGTGTGT---CAAGCCAGCCCTCAGATAGTCAAGCCAGAGCTTCAC	513
Db	1979	oHisGlyLysLeuSerHisPheGlyProGlyProIlnbLeuGlySerProGlnSerThrAs	1999
OY	512	ATATGTGTCTTTGTTGGATCAGGAGCACCACTTGCTCCATCACTCTTGAAGAACAG	453
Db	1999	nTyThrValAlaThrGlyAsn-----PheHisPro-----Se	2010
OY	452	CTGCCAGCCTGGGGGCGAGAGTGAAGAGTAACAGGCGTGAATGTTCTCTTTGTGAGG	393
Db	2010	rGlySerProLeuGlyProSerSerGlySerThrGlyLysSerTyGlyLeuSer-ProL	2030
OY	392	CAAAACACTTCTCCCTATTTGACTGCTTGAGAACACAGTATACCTCTCTCACTACCT	333
Db	2030	euaArgProProSerValLeu-----	2036
OY	332	CAGCACCCCTCTACTCTCCAGTGGGGATCACTATCTATCACTCCAGAGAGTCTCTG	273
Db	2037	----ProProProAlaProAsp---GlySerLeuProTyL---LeuSerHisGlyAlas	2053
OY	272	CCATCTCCCTTGAAGGGCTTCTCCACT-----G	243
Db	2053	erGlnArgSerGlyLysThrSerProValGluLysArgGluAspProGlyThrGlyMetG	2073
OY	242	GCTCAACCTGTCAGAGTGTGTGTTGGAGCACCAAGAGCCGACACTGGGCAAGGGCTCT	183
Db	2073	lySerSerLeuAlaThrValaGluLeuProGlyThrGlnAspProGlyLysSerGlyLeuS	2093

QY. 182 CACAGACATCCACAGCCTCAGCCAGCTTGACACTGAAAGAGGACAGCTC 133  
||||| ||| |||::|||  
Db 2093 ergIntThrCluLeuGluGlyGlnArgGlnArgGlnArgLeuArgGlnLeu 2109

## RESULT 2

ALR protein - human

C:\Date: 24-Mar-1999 #sequence\_revision 24-Mar-1999 #text\_change 21-Jul-2000

R; Prasad, R.; Zhadanov, A.B.; Sedkov, Y.; Bullrich, F.; Druck, T.; Rallapalli, R.; Yano

A;Title: Structure and expression pattern of human ALR,

A;Accession: T03454

A: Molecule type: mRNA

A:Cross-references: EMBL:AF010403. NTD:G2358284. PTN:AA051734 1. PTD:G2358285  
A;RESIDUES: 1-3262 <PKA>

C;Genetics:  
A:Genetics: ATP

A;Map position: 12

C;Keywords: alternative splicing

Alignment Scores:

Pred. No.:	0.0291	Length:	5628
Score:	127.50	Matches:	106
Percent Similarity:	33.97%	Conservative:	36
Best Local Similarity:	25.33%	Mismatches:	149
Query Match:	4.98%	Indels:	128
DB:	2	Gaps:	22

US-09-989-919-15 (1-1397) x T03454 (1-5262)

1219 CTTTCATTGGAAGCAATGAG-----GCAGAAAGGCCCGGCC 1181

Db 2070 LeuGlyLeuArgProGlnGluProProProAlaGlnAlaLeuAlaProSerProProSer 2089

1180 CACCCC-----CGCCCAGGTCCTGGCTCAAACACAC-----CTGCTC 1142

Db 2090 HisProaspIlepheargProGlyserTyrrThraspProTyralaGlnProProLeuThr 2109

1141 CCTGACCCCACTCTTGGCTCATGCTGAGGTGTGCACCTCTGCCCT----- 1096

Db 2110 ProArqProGln-ProProProProGluSerCysAlaLeuProProArqSerLeuPr 21299

1095 -----GACCCCTGGGCTGGCCTGGGAGTGTCCCCTGTCCGGAGGTCAGAGATAGCCTCCCC 1040

```

Db      2129 oSerAspProphSerArqValpöValSerproGInSerGInSerSerGInSerPr 2149
          |||||:::|::| |||||:::| |||||

```

1039 AAGTACAGATAAACAAATCAAGGACCTAATCCTCCACACGACCTAAC GGG

$\frac{1}{2}$

[illegible][illegible][illegible]

.....

[illegible]

2130 gftocnflotocluvaimafucy8-----Alaalyseleuamablsll 2212

[illegible]



Db 2212 rSerLeuGlyAlaGlyPheProAlaAlaLeuProAlaGlyProAlaGlyLeu-- 2231  
 QY 707 ATTTAAATTTGGCATTTATTTGACAGCTTTGCTGTCGATCATCTGTCTGGAGATGT 648  
 Db 2232 -----HisAlaValSerAlaProSerGlyGlnProProAlaPheValArgSerProGlyTh 2249  
 QY 647 AGCTTTGTCTGACAGGGGTTCCAGATGACATGACAGTCTGCTCCAGTGCAGTCACT----- 593  
 Db 2249 rGlyAlaPheValGlyThrProSer-----PromethArgPheThrPhePhe 2264  
 QY 592 -----CCCTCCAGAGGCAAGCCAGCTG 570  
 Db 2264 oGlnAlaValAlaGlyLeuProSerLeuLysProProValProGlnProGlyLeuProProP 2284  
 QY 569 CTGAGGATAGTACAGCTGCT---CAGCCAGACCTCCAGATAGTACAGCAAGGCTTCCAC 513  
 Db 2284 oHisGlyIleAsnSerHisPheGlyProGlyProThrLeuGlyLysProGlnSerThrAs 2304  
 QY 512 ATATGTGCTTTGTGGATCAGGAGACCACTTGCTTCATCCAGCTTTGAAAAAGACAG 453  
 Db 2304 nTyThrValAlaThrGlyAsn-----PheHisPro-----Se 2315  
 QY 452 CTGCCACCTGGGGGGGAGAGGTGAGGTAAACAGGCTGATGTTCTTTCTTGAGG 333  
 Db 2315 rGlySerProLeuGlyProSerSerGlySerThrGlyLysSerTyGlyLeuSer-ProL 2335  
 QY 392 CAGACCATCTCTCCATTTGACTGCTTGAAGACAGTATACCTCTCTCATACCT 333  
 Db 2335 euArgProProSerValLeu----- 2341  
 QY 332 CAGACCCCTCTGACTTCCGATGGGGATCACTATATCACTCCAGAGGCTCTG 273  
 Db 2342 -----ProProProAlaProAsp---GlySerLeuProTyL---LeuSerHisGlyAlaAs 2358  
 QY 272 CCAATCCCTTGAAGGGCTTCTCCACT-----G 243  
 Db 2358 eArgInArgSerGlyIleThrSerProValGlyLysArgLysAspProGlyThrGlyMetG 2378  
 QY 242 GCTCACCCTGCGAGTGTCTGGTGTGACACAAAGGCCGACACTGGGCACTGGCTT 183  
 Db 2378 LysSerSerLeuAlaThrAlaGlyLeuProGlyThrGlnAspProGlyMetSerGlyLeuS 2398  
 QY 182 CACAGACATCCAGACGCTCAGCCAGCTTGAACACTGAAGAGGAGCACTC 133  
 Db 2398 eArgInThrGlyLeuGlyLysGlnArgGlnArgLysArgLysGlnLeu 2414

## RESULT 3

149705  
 glutamate receptor channel subunit epsilon 3 - mouse  
 C:Species: Mus musculus (house mouse)  
 C:Date: 02-Aug-1996 #sequence\_revision 02-Aug-1996 #text\_change 21-Jan-2000  
 C:Accession: I49705  
 R:Kutsuwada, T.; Kaahiwabuchi, N.; Mori, H.; Sakimura, K.; Kushiya, E.; Araki, K.; Meguri  
 Nature 359, 36-41, 1992  
 A:Title: Molecular diversity of the NMDA receptor channel.  
 A:Reference number: 149704; PMID:92310564; PMID:1377365  
 A:Accession: I49705  
 A:Status: preliminary; translated from GB/EMBL/DBD  
 A:Molecule type: mRNA  
 A:Residues: 1-1239 <RES>  
 A:Cross-references: GB:D10694; NID:G538239; PIDN:BA001536.1; PID:G538240  
 C:Superfamily: N-methyl-D-aspartate receptor 2C; glutamate receptor homology  
 C:Keywords: neurotransmitter receptor  
 F:425-852/Domain: glutamate receptor homology <GRH>

## Alignment Scores:

Pred. No.: 0.0388 Length: 1239  
 Score: 126.00 Matches: 94  
 Percent Similarity: 30.61% Conservative: 23  
 Best Local Similarity: 24.61% Mismatches: 112  
 Query Match: 4.92% Indels: 153  
 DB: 2 Gaps: 24

US-09-989-919-15 (1-1397) x I49705 (1-1239)  
 QY 1196 AGAAGAGCCCGGCGCCCGCCAGGCTCTGATCAACACAGCTGCTCCCTTA 1137  
 Db 927 ArgAlaAlaProAlaPro-----ThrThrSerGlyPro----- 937  
 QY 1136 CCCCAGCTTGGCTCAATGCTAGGTGACACCTGCTCCCTGACCCCTGGG----- 1086  
 Db 938 -----ArgSerGlyThrProGlyPro-----ProGlyGlnProSer 949  
 QY 1085 ---CTGGCTGGAGATGCTCCCTGTGGAGGTCAAGATAGCTTCCAGATACGAT 1029  
 Db 950 ProSerGlyITrArgProPro-----GlyGlyGlyArg---ThrProLeuAlaArgAG 966  
 QY 1028 CAGCCACATCTGGAGACATCCCGCCGCAAG-----TCTCTCTCCAGACTTAAGT 978  
 Db 967 AlaProGlnProProAlaArgProAlaThrCysValaGlySerProGlnProAlaValSer 986  
 QY 977 GAGAGAACATCCACAGATGATGTGCCACGAGCCCGCCGCTGTGTGSCATGTG 918  
 Db 987 ArgAlaSerCysArgHisAlaTrpAspAlaArg-----TrpProVal 1000  
 QY 917 CAGCAGCAGCAGCAGGCGCTTAACTACATTTCAACATTAATTAACACAGTATGGCGCT 858  
 Db 1001 ArgValaGlyHisGlnGly-----SerHisLeuSerAlaAsp 1012  
 QY 857 GCCCAGAGAACCGTCCAGAGGAGGAGCAGAGCCCGCCAGTGGGAGACTAGCTGACTTG 798  
 Db 1013 GlnArgArgAlaLeuProGlnArg-----SerLeu 1022  
 QY 797 CTTCAGCCCGAC-----AGGGTTGAGAGGAGCAGTC 768  
 Db 1023 LeuHisAlaHisLeuGlyHisIleTySerSerPheProArgAlaGlnArgSerGlyArgProPhe 1042  
 QY 767 ATACCCCGCATGATTTCTGACTAGAGCCAGCA-----GTCCCATAGCTAAGC 720  
 Db 1043 LeuPro-----LeuPheProGlnProProGlnProAlaAspArgLeuProLeuLeuGly 1059  
 QY 719 TGGTAACATCAATTTAATTTGACATTTATGACAGCTTTGCTCGTGCACATCTGTGT 660  
 Db 1060 ---ProGlnGlnLeuAlaArgArgGlyAlaLeuLeuAlaAlaTrpAla----- 1075  
 QY 659 GTCTGGAGATGATTAAGCTTTGTCTGACAGGGTTCACAG---TGACATGACAGTGTGCACT 603  
 Db 1076 -----ArgGly-ProArgProArgHisAlaSerLeuProSe 1087  
 QY 602 GCAAGTC-----ACTCCCTCCACAGGCAAGCCAGC---TGCTGAGGATAGTC 558  
 Db 1087 rSerValAlaGlyAlaPheThrArgSerAsnProLeuProAlaArgCysThrGlyHisAl 1107  
 QY 557 AGCTGTGACGCGAGCGCTCAGATAGGTGACGACAGAGGCTTCACATATGTGTCTTGT 498  
 Db 1107 AcCysAlaCysProCysProGln----- 1114  
 QY 497 GGGATCAGGAGACCACTTGCTGCATCAGCTTTGAAAAAGACAGCTGCGCA----- 446  
 Db 1115 -----SerArgProSerCysArgHisValaAl 1123  
 QY 445 -----CTGGGGGGCAGAGGTGAGGTAAACAGGCTGATGT 408  
 Db 1123 aGlnThrGlnSerLeuArgLeuProSerTyArgGlyAlaCysValaGlyGlyAlaProAl 1143  
 QY 407 TTCTCTTTGTGAGGAGGAACCATTTCTCCCTAATTGACTGCTTGAAGACAGTATAC 348  
 Db 1143 aGlyValAlaAla-----ThrTrpGlnPhe 1151  
 QY 347 TCTCTCTCAT-----CACCTCAGACCCCTGACTTCCGATGGGGGATCACTAT 297  
 Db 1151 oArgGlnHisValaCysLeuHisIleThrHisIleuProPheCysTrpGlyThrValaCys 1171  
 QY 296 CTATCAACCTCCAGAGAGGTCTGTCATCTCCCTTGAAGGCGCTTCCCATGCTGCTAC 237  
 Db 1171 sArgHisProProProCysSerSerHisSerProTrp----- 1183

QY 236 CTGTCCAGTGTGCTGTGGT-----GACCACAAAGCCCGCACTGGG 192  
 Db 1184 -----LeuilegilyThrtPrGluProProSerHisArgIlyrghThleuGI 1199  
 QY 191 CARG 188  
 Db 1199 yLeu 1200

RESULT 4  
GNMVR4

structural polyprotein - rubella virus (strain RA27/3 vaccine)  
 N:Contains: capsid protein; membrane glycoprotein E1; membrane glycoprotein E2  
 C:Species: rubella virus  
 C>Date: 30-Sep-1990 #sequence\_revision 30-Sep-1990 #text\_change 16-Jul-1999  
 C/Accession: S04800  
 R:Nakhaei, H. L.; Thomas, D.; Zheng, D.; Liu, T. Y.  
 Nucleic Acids Res. 17, 4393-4394, 1989  
 A:Title: Nucleotide sequence of capsid, E2 and E1 protein genes of rubella virus vaccine  
 A:Reference number: S04800; PMID:89296505; PMID:2740235  
 A:Accession: S04800  
 A:Molecule type: mRNA  
 A:Residues: 1-1063 <MAX>  
 A:Cross-references: GB:X14871; NID:961915; PIDN:CAA33016.1; PID:961916.  
 A>Note: the authors translated the codon CGC for residue 207 as Val and AGC for residue  
 C:Superfamily: rubella virus structural polyprotein  
 C:KeyWords: capsid protein; glycoprotein; polyprotein; transmembrane protein  
 F:1-300/Product: capsid protein C #status predicted <NPC>  
 F:278-598/Domain: transmembrane #status predicted <TM1>  
 F:301-582/Product: membrane glycoprotein E2 #status predicted <TM2>  
 F:559-555/Domain: transmembrane #status predicted <TM2>  
 F:583-1063/Product: membrane glycoprotein E1 #status predicted <E1G>  
 F:1032-1050/Domain: transmembrane #status predicted <TM3>  
 F:353,371,410,429,658,759,791/Binding site: carbohydrate (Asn) (covalent) #status predicted

## Alignment Scores:

Pred. No.: 0.131 Length: 1063  
 Score: 119.50 Matches: 106  
 Percent Similarity: 31.33% Conservative: 35  
 Best Local Similarity: 23.56% Mismatches: 143  
 Query Match: 4.67% Indels: 166  
 DB: 1 Gaps: 24

US-09-989-919-15 (1-1397) x GNMVRA (1-1063)

QY 1184 GCGCCACCGCCGCGCCAGCTCTGCTCAAC-----ACACTGCTCCCTGACCC 1134  
 Db 77 AlArProProProGluGluArgIlnGlnUserArgSerGlnThrProAlaProLyPro 96  
 QY 1133 CAGCTTGGCTCATGCTGAGGTGCGACCTTGCCCTGACCCCTGGGGTGGCTGGAG 1074  
 Db 97 SerArg-----AlaPro--ProGlnGlnProGln-- 105  
 QY 1073 TGTCCCTGTCGGAGAGTCAGATAGCTCCCGAGTACAG-----AAT 1029  
 Db 106 -----ProProArgMetGlnThrArgIlyrghIlySer 116  
 QY 1028 CACCCACATCTCGAGAGATCCCGCCAGTCTCTGTCAGACCTAGAGAGA--- 972  
 Db 117 AlArProArProGluLeuGlyProProThrAsnProPheGlnAlaAlaValAlaArgGly 136  
 QY 971 -----ACACTCCACAGATGATGTGCCCAAGCAGCCCGCCGCTGCTGT----- 926  
 Db 137 LeuArProProLeuHisAspPro--AspThrGln--AlaProThrGlnAlaCysValTh 155  
 QY 925 -----GGCATGTGCGAGAGCAGACAGAGGGCTTTAGCTACACTTCACACTATTGAACAC 870  
 Db 155 rSerTrpLeuTrpSerGluGlyGlyAlaValAlaPheThyArgValAspLeuHisPheTh 175  
 QY 869 AGTATGGGGGTGCGCCACGGAACCGTCCAGAG----- 836  
 Db 175 rAsnLeuGlyThrProProLeuAspGluAspGlyArgTrpAspProAlaLeuMetTyAs 195

QY 835 -----GGGGACAGCCCGCCAGTGGGGAGACTAGACCTGAGCTGTGTTACAGCCACAGG 783  
 Db 195 nProCysGlyProGluProPro--AlaHisValAlaArgAlaTyAsnGlnProAlaGI 214  
 QY 782 GTTGGAGAGCAGATATACCCCCAGATGATCTCGATAGAGCCAGACAGTCCCATAGCTA 723  
 Db 214 yAspValAlaArgValAlaTrpGlyLeuGlyGlnUArgThrTyAlaGlnGlnAspPheArgVa 234  
 QY 722 GCGGTGAACATCACTTTTAACATTGGACATTATGACGCTTTGTCTGTCGACATCTG 663  
 Db 234 lGlyGly-----ThrArgTrpHisArgLeuLeuArgMetProValAlaGlyLy----- 249  
 QY 662 TCTGTCTGGAGATTAGCT-----TTCCTCCAGAGGTTCCAGATGACATGACATGCTGCC 606  
 Db 250 -LeuArgGlyAspThrAlaProLeuSerPro-----HisThrThrGlnUArg 264  
 QY 605 AGTGCAGATCACTCCCTCCACAGAGAGCCAGCTGCTGAGATAGTACAGCTGCAGCC 546  
 Db 264 gIleGlu--ThrArgSerAlaArgHisProSer----- 274  
 QY 545 GAGCCCTCAGATAGTCAAGCCAGAGCCCTTTCACATATGTGCTTGTGGATCAGAGAC 486  
 Db 275 -SerIleArgPheGlyAlaProGlnAlaPheLeuAlaGlyLeuLeuAlaAlaValAl 294  
 QY 485 CACTTGGCTCCATCCAGTCTTGAAAAAGACAGCTGCCAGCTGGGGGGGACAGGTGAGA 426  
 Db 294 a-ValIlyThrAlaArgAlaGlyLeuGlnProArgValAspMet----- 308  
 QY 425 GGTACAGAGGCTGAGATGTTCTCTTCTTGGAGGACAAAGACATTCCTCAATTGACTGC 366  
 Db 308 ----- 308  
 QY 365 CTTGAAGACAGATATACCTCTCTTCATACCTCAGACACCCCTGCACTTCCGATGGG 306  
 Db 309 -----AlaAlaProProMet----- 313  
 QY 305 GATCACTATCTATCAACCTCCAGAGAGTCTGCCATCTCCCTTGGAAAGGCTTCCCA 246  
 Db 314 -----Pro-ProGlnPro-----Pro 318  
 QY 245 CTGGCTCACTGTCCAGATGTCTGTGTGTGATGACCAAAAGGCCAGACTGGGCATGGC 186  
 Db 319 ArgAlaHisGlyGlnHisTyArgIlyHisHisHisGlnLeuProPheLeuGlnHisAsp 338  
 QY 185 TCTCAACAGACATCCACAGCTCAGCCAGGTGAACACTGAAGAGAGGAGCTCCCTGGT 126  
 Db 339 gAllyHisHis-----GlyGlyThrLeuArgVal 347  
 QY 125 GGTAGAGATGAGCCAGAGCGGTAGT-----CTTCCCTGGGGATGG 87  
 Db 348 GlyGlnHisHisArgAsnAlaSerArgValLeuProGlyHisIleTrpLeuGlnGlyTyTr 367  
 QY 86 TGCTGTCTGGGA-----TACACTGTACTCGGTACTGC----- 54  
 Db 368 GlyCysTyAsnLeuSerAspTrpHisGlnGlyThrHisValCysHisThrIlyHisMet 387  
 QY 53 -----TGCTTG-----CCGTGAGCTTGTGACAGAT 30  
 Db 388 AspPheTrpCysValGlnHisAspArgProProProAlaThrProHisIleLeuThrThr 407  
 QY 29 ACTGCCCGCTCCGGTACAGGTGCA 6  
 Db 408 AlaAlaHisSerThrThrAlaAla 415

## RESULT 5

## GNM777

structural polyprotein - rubella virus (strain HPV77)  
 N:Contains: capsid protein; membrane glycoprotein E1; membrane glycoprotein E2  
 C:Species: rubella virus  
 C>Date: 30-Jun-1991 #sequence\_revision 30-Jun-1991 #text\_change 16-Jul-1999  
 C/Accession: J00087  
 R:Zheng, D.; Dickens, L.; Liu, T. Y.; Nakhaei, H. L.  
 Gene 82, 343-349, 1989

A:Title: Nucleotide sequence of the 24S subgenomic messenger RNA of a vaccine strain (HF  
 A:Reference number: JQ0087; MUID:90060825; PMID:2583526  
 A:Accession: JQ0087  
 A:Molecule type: mRNA  
 A:Residues: 1-1063 <ZHE>  
 A:Cross-references: GB:M30776; NID:G333974; PIDN:AAA47421.1; PID:G333975  
 C:Superfamily: rubella virus structural polypeptide  
 C:Keywords: capsid protein; glycoprotein; polypeptide; transmembrane protein  
 F:1-300/Product: capsid protein #status predicted <CAP>  
 F:278-298/Domain: transmembrane #status predicted <TM1>  
 F:301-582/Product: membrane glycoprotein E2 #status predicted <E2>  
 F:539-555/Domain: transmembrane #status predicted <TM2>  
 F:583-1063/Product: membrane glycoprotein E1 #status predicted <E1>  
 F:1032-1050/Domain: transmembrane #status predicted <TM3>  
 F:353,371,429,658,759,791/Binding site: carbohydrate (Asn) (covalent) #status predicted

## Alignment Scores:

Pred. No.:	0.144	Length:	1063
Score:	119.00	Matches:	98
Percent Similarity:	32.46%	Conservative:	39
Best Local Similarity:	23.22%	Mismatches:	127
Query Match:	4.65%	Indels:	159
DB:	1	Gaps:	22

US-09-989-919-15 (1-1397) x GNMV77 (1-1063)

QY 1184 GCCCACACCCCGCCGAGTCTGCTCAAC-----ACACTGCTCCCTGACCC 1134  
 |||||  
 Db 77 AlAPrProPrProGlulnArglnGlnSerArgSerGlnThrProAlaProLysPro 96  
 QY 1133 CAGCTTGGCTCATGCTGAGGTGTCACCTGCGCCCTGAGCCCTGGGCTGGGAG 1074  
 |||||  
 Db 97 SerArg-----AlaPro-----ProGlnInProGln----- 105  
 QY 1073 TGTCCCTGTGGGAGGTACAGATAGCTCCCGAGTAAAG-----AAT 1029  
 |||||  
 Db 106 -----ProPrArgMetGlnThrGlnArgGlnGlySer 116  
 QY 1028 CACCCACATCTGAGAGATCCCGCCAGTCTGCTGTCAGACCTAAGCTGAGAG-- 972  
 |||||  
 Db 117 AlAPrProPrProGlulnArglnGlnSerArgSerGlnThrProAlaProLysPro 136  
 QY 971 -----ACACTCAAGATGATGTCACCGCCGAGCCGCTGCTGT----- 926  
 |||||  
 Db 137 LeuArgProPrProLeuHisAspPro--AspThrGln-AlaProThrGlnAlaCysValIn 155  
 QY 925 -----GGCATGTGACGACGACGACGAGGGCTTACCTACATTTCTACACTATTGAAC 870  
 |||||  
 Db 155 rSerTrpLeuTrpSerGlnGlnGlnAlaValPheTyArgValAspLeuHisPheTh 175  
 QY 869 AGTATGAGGGCTGCCACGGAACCGTCCAGGGA----- 836  
 |||||  
 Db 175 rAnleuGlnThrProPrProLeuAspGlnAspGlnArgTrpAspProAlaLeuMetCys 195  
 QY 835 -----GGGACACAGCCCGCCAGTGGGAGACTAGCTTGTGTTACAGCCACAG 783  
 |||||  
 Db 195 nProCysGlnProGlnProPro--AlaHisValAlaArgAlaLysAsnInProAlaG 214  
 QY 782 GTTGGAGAGCAGTATACCCCATGATTTCTGACTAGAGCCAGCAGAGTCCCATAGCTA 723  
 |||||  
 Db 214 yAspValArg-----GlyValTrpGlnGlnGlnArgThrTy--Ala- 228  
 QY 722 GCGTGTAAACTCAATTTAAGATT-----TGGCATTTATGACCTTTGTC 675  
 |||||  
 Db 229 -----GlnGlnAspPheArgValGlnGlnArgTrpHisArgLeuLeuArgMet 246  
 QY 674 TGGTCAATCTGTCTGTGGAGGTGTAGCTTGTGCTGAGGGGTCCCGAGTACATGC 615  
 |||||  
 Db 246 roVal-----ArgGlnLeuAspGlnAsp----- 253  
 QY 614 AGTGTCTGCCAGTCAATCTCCCTCCACAGGAAAGCCAGCTGTGAGAGATAG--TCAG 556  
 |||||  
 Db 254 -----ThrAlaProLeuProPrHisThrThrGlnArgGlnGlnThrArgSerA 270

QY 555 CTGTGACGCG--AGCCCTCAGATAGTGTACGACGAGGCTTACATATGTGCTGT 499  
 |||||  
 Db 270 lArArgPrProTrpAlaGlnLeuArgPheGlnAlaProGlnAlaPheLeuAlaGlnLeu 290  
 QY 498 TGGATCAGAGGACCACTGTGCTCCATCCAGTCTTGAAGAAAGACCACTGGCAGCTGG 439  
 |||||  
 Db 290 euAlaAlaValAla-ValGlnThrAlaArgAlaGlnLeuGlnProArgAlaAspMet-- 308  
 QY 438 GCGAGAGGTGAGAGGTAAACAGGCTGGATGTTCTTCTTGTGAGGCAACCATTTCTC 379  
 |||||  
 Db 308 ----- 308  
 QY 378 CTAATTGACGCTTGAAGACAGATACCTCTCTGTCACCTGACGACCCCTCTG 319  
 |||||  
 Db 309 -----AlaAlaProPheMet 313  
 QY 318 ACTTCGATGGGGATCACTATCTATCAACCTCCAGAGAGTCTCCCATCTCCCTGGA 259  
 |||||  
 Db 314 -----Pro-ProGlnPro----- 317  
 QY 258 AGGACTTTCCTCCACTGCTCCTGCTCCAGGCTGTGCTGTGTCACCAAGGCCCA 199  
 |||||  
 Db 318 -----ProArgAlaHisGlnGlnHisTyArgGlnHisHisGlnInLeuProPr 334  
 QY 198 CACTGGCAGTGGCTTCACAGACATCCACAGCCTCAGCCAGTTGAACCTGAAGAGAG 139  
 |||||  
 Db 334 leuGlnHisAspGlnHis-----GlnGln 343  
 QY 138 CAGCTCCGCTGTGTAGAGTGGCCAGCGTAGT-----CT 100  
 |||||  
 Db 343 yThrLeuArgValGlnGlnHisArgAsnAlaSerAspValLeuProGlnHisTyPhe 363  
 QY 99 TCTTGGGAGTGTCTGTCTGGA-----TACACTGGTACTCGTACTGC 54  
 |||||  
 Db 363 uGlnGlnGlnTrpGlnCysTyArgAsnLeuSerAspTrpHisGlnGlnHisValCys 382

## RESULT 6

A39255

cytokine receptor common beta chain precursor - human

C/Species: Homo sapiens (man)

C/Date: 30-Jun-1993 #sequence\_revision 30-Jun-1993 #text\_change 02-Sep-1997

C/Accession: A39255

R:Hayashida, K.; Kitamura, T.; Gorman, D.M.; Arai, K.; Yokota, T.; Miyajima, A.

A:Title: Molecular cloning of a second subunit of the receptor for human granulocyte-ma-

A:Reference number: A39255; MUID:91088571; PMID:1702217

A:Accession: A39255

A:Molecule type: mRNA

A:Residues: 1-897 &lt;HAV&gt;

A:Cross-references: GB:M30275

C/Comment: The human high-affinity IL-3, IL-5, and GM-CSF receptors have ligand-specific

C/Genetics:

A:Gene: CSF2RB

A:Cross-references: GDB:126838; OMIM:138981

A:Map position: 22q13.1-22q13.1

C:Superfamily: interleukin-3 receptor beta chain; cytokine receptor homology

C:Keywords: alternative splicing; cytokine receptor; duplication; transmembrane protein

F:1-16/Domain: signal sequence #status predicted &lt;Sig&gt;

F:17-897/Product: cytokine receptor common beta chain #status predicted &lt;Ext&gt;

F:17-443/Domain: extracellular #status predicted &lt;Ext&gt;

F:35-232/Domain: cytokine receptor homology &lt;CRS1&gt;

F:250-431/Domain: cytokine receptor homology &lt;CRS2&gt;

F:444-460/Domain: transmembrane #status predicted &lt;TM&gt;

F:461-897/Domain: intracellular #status predicted &lt;INT&gt;

## Alignment Scores:

Pred. No.:	0.173	Length:	897
Score:	118.00 <td>Matches:</td> <td>104</td>	Matches:	104
Percent Similarity:	34.37% <td>Conservative:</td> <td>51</td>	Conservative:	51
Best Local Similarity:	23.06% <td>Mismatches:</td> <td>152</td>	Mismatches:	152
Query Match:	4.61% <td>Indels:</td> <td>144</td>	Indels:	144
DB:	1	Gaps:	22

US-09-989-919-15 (1-1397) x A39255 (1-897)

QY 1221 GGCTTCATTGAAAGCAATAGAGCAAGAAAGCCCGGCGCCACCCCGCCCGAGGTCTT 1162  
 DB 464 GYIATYGLYIYRARGLEUARGLYETRGLULULYSILEPROANPROSERLYS 483  
 QY 1161 GGC-----TCMAACACACCTGCTCCCTGACCCCACTTGGCTCATGCTAGAGG 1111  
 DB 484 SerHisLeuPheGlnSerHisValGlnLeuTyrProGlnTyrSerMetSerAlaPhe 503  
 QY 1110 TGCACCTGCGCCCT-----GACCCCTGGGCTGCGCTG-----GGA 1075  
 DB 504 ThrSerGlySerProHisGlnGlyProTyrGlySerArgPheProGlnLeuGlnGly 523  
 QY 1074 GTGTCCCTGTC-----GGAGAGTCAGAGATAGCCTCCCGAGTACAGAAATCACCCACA 1021  
 DB 524 ValPheProValGlyPheGlyAspSerGlnVal-----SerProLeu 537  
 QY 1020 TCCTGAGACATCCCGCCCAAGTCTCTGTCACAGCTTAAGCTGAGAGAACTCCACGA 961  
 DB 538 ThrIleGlnAspProLysHisValCysAspProProSerGlyProAspThrThrProAla 557  
 QY 960 TGGATT-----GTCCCGCAG-----CAGCCCGCACCGCCTGCTGTCGCA 922  
 DB 558 AlaSerAspLeuProThrGlnGlnProProSerProGlnProGlnProAlaAlaSer 577  
 QY 921 TGTGACACACAGACAGAGGCGCTTAGCTA----- 891  
 DB 578 HisThrProGlnGlnAlaSerSerPheAspPheAsnGlyProTyrIleGlnGlyProP 597  
 QY 890 -CATTCACACATTTGAACACAGTATGGGGCTGCCAGGAACTCCCGAGGAGGG 832  
 DB 597 OHAsSerArgSerLeuProAspIleLeuGlnGlnProGlnProGlnGlnGlyGlyS 617  
 QY 831 CACAG-----CCCCAGTGGAGAGATAGCTGAGCTTACAGCCACCAAGGGTGG 778  
 DB 617 erGlnLysSerProProGlnGlnSerIleuGlnIlyrLeuGlySerLeu---ProAlaGly 636  
 QY 777 AGAGGACATACACCCCATGATTCCTGAGCTAGAGACCAAGCTCCCATAGCTAGAGCTG 718  
 DB 636 InValGlnLeuValPro----- 641  
 QY 717 GTAACCTCACATTTTAACATTGGCATTTATGACGTTTGTCTGTCATCTGTCTGT 658  
 DB 641 ----- 641  
 QY 657 CTGGAGTGTAGCTTGTCTGACAGGGTCCAGAGTACATGACAGTGCCTCCAGTGCAG 598  
 DB 642 -----LeuAlaGlnAlaMetGlyProGlnGln-----AlaValGlnValGln 656  
 QY 597 TCACCTCCCTCCACAGGAGCCAGCTGCTGAGAGATGACAGTGTGTCACCCAGCCCTC 538  
 DB 656 rGArgProSerGlnGly-----AlaAlaGlySerProSerLeuG 669  
 QY 537 AGATAGTTCAGCAGAGAGCCTTACATATGTGTCTTGTGGATTCAGGAGACACTTGGC 478  
 DB 669 IuSerGlyGly-----GlyProAlaProProAla 679  
 QY 477 TCATCAGCTTGAAGAAAGACAGCTGCGACCTGGGGGAGAGAGTGAAGTGAACAG 418  
 DB 679 euGlyPro-----ArgValGlyGlyGlnAspGlnLysAspSerP 692  
 QY 417 GGTGATGATTTCTTCTTGGAGCAAGACCATTTCTCCATATTGACTGCTTGAGA 358  
 DB 692 roValAlaIleProMetSerSerGlyAspThrGlnAspProGlyValAlaSerGlyTyr 712  
 QY 357 CAGCTATACCTCTCTCATCAGCTCAGACACCCCTCTGACTTCGATGGGAGTACTA 298  
 DB 712 alSer-SerAlaAspLeuVal-----PheThrProAsnSerGlyAlaSer 726  
 QY 297 TCTATCAAGCTC-----CAGAGAGGTCTGCGCATCTCCCTT 262

DB 727 SerValSerLeuValProSerIleuGlyLeuProSerAspGlnThrProSerLeuGlyPro 746  
 QY 261 GGAAGGCTTCT---CCCACTGGCTCAGCTGTCACAGTG----- 226  
 DB 747 GlyLeuAlaSerGlyProProGlnValAlaProGlyProValLysSerGlyPheGlnGlyTyr 766  
 QY 225 GTCTGGTGGTGAACCAAGAGCGGACAGCTGGGATGGCTCTCACAGACATCCACAGCC 166  
 DB 767 ValGlnLeuProProIleGlnGlyTyr-SerProArgSerProAlaGlnAsnProValP 786  
 QY 165 T-----CAGCCAGTTGAACACTGAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 130  
 DB 786 CProGlnAlaLysSerProValLeuAsnProGly-----GlnArg 799  
 QY 129 TGGTGTAGATGGCCAGCAGCGGTAAGCTTCTTGGGGATGATGCTGTCTGGATACAC 70  
 DB 799 gProAlaAspValSerProThrSerProGlnProGlnGlyLeuValLeuGlnGlnVal 819  
 QY 69 TGTACTCGGTACTGCTGCTGCTGCGG 45  
 DB 819 IGLYASP---TYRCYSPHELEUPRO 826

## RESULT 7

T45025  
 mucin MUC5B, tracheobronchial [imported] - human (fragment)  
 C/Species: Homo sapiens (man)  
 C/Date: 21-Jan-2000 #sequence\_revision 21-Jan-2000 #ext\_change 21-Jul-2000  
 C/Accession: T45025  
 R/Dessey, J.L.; Guyonnet-Duperrat, V.; Porchet, N.; Aubert, J.P.; Laine, A.  
 J. Biol. Chem. 272, 3168-3178, 1997  
 A>Title: Human mucin gene MUC5B, the 10.7 kb large central exon encodes various alternat  
 A/Reference number: Z22899; MUID:97166151; PMID:9013550  
 A/Accession: T45025  
 A/Status: preliminary; translated from GB/EMBL/DBJ  
 A/Molecule type: DNA  
 A/Residues: 1-3570 <DBS>  
 A/Cross-references: EMBL:Z72496; NID:G1834502; PIDN:CAA6577.1; PID:G1834503  
 A/Experimental source: placenta  
 C/Genetics:  
 A/Gene: MUC5B

## Alignment Scores:

Pred. No.: 0.172 Length: 3570  
 Score: 118.00 Matches: 100  
 Percent Similarity: 31.768 Conservative: 55  
 Best Local Similarity: 20.498 Mismatches: 159  
 Query Match: 4.61% Indels: 174  
 DB: 2 Gaps: 19

US-09-989-919-15 (1-1397) x T45025 (1-3570)

QY 1294 CCAACAGAGTACAAAGATTGTGATATTCAACTCGAGAGAGGAGAAAGCTGTGTT 1235  
 DB 795 ProSerSerThrGlnThrSerGlyThrProProSerLeuThrThrThrAlaThrThr 814  
 QY 1234 TTGGCTGCTTTAGAGCTTTCATTGAAGCAATAGGAGCAAGAGCCCGGCCACCC 1175  
 DB 815 ILeThrAla-----ThrLysThrThrAsnProSerThrProGlyThrThrThr 832  
 QY 1174 CGCCCGCTGCTGAGCTCAACACACACTGCTCCCTGACCCGAGTGGCTCATAGCTGA 1115  
 DB 832 oIlePro-----ProValLeuThr----- 838  
 QY 1114 GGTGTCACCTCTGCCCCCTGACCCCTGGGCTGGCGTGGAGAGTCCCTGTGGGAGGTC 1055  
 DB 839 -----ThrThrAla-----ThrThrProAlaAlaThrSe 848  
 QY 1054 AGAGATAGCTCCCGCAGTACAGAAATCACACATCATCTGGAGATCCCGCAGAGTCTC 995  
 DB 848 rSerThrValThrProSerSerAlaLeuGlyThrThrHsrThrProPro----- 864  
 QY 994 CTGTCCAGACTTAAGCTGAGAGAAACATCAGATGATGTCCCGACAGCCCGCACACC 935

Db 865 -ValProaenThrThraLathrThrhISglYargSerLeuSerProSerSerProhiSth 884  
 QY 934 GCCTGCTGTGACATGTCACAGCAGCAGGCGGCTTGTAGTACATTTCACACATATTG 875  
 Db 884 rValcYsthrAlaThrThrSerAlaThrSerGlyIleLeuGlyThr-ThrhIleIethc 904  
 QY 874 AACACAGTATGGGCTGCCAC----- 852  
 Db 904 lProSerThrGlyThrSerHisThrProAlaAlaThrThrGlyThrGlnHisSer 924  
 QY 852 ----- 852  
 Db 924 hrProAlaLeuSerSerProhiSerSerSerAlaThrThrGlyThrGlySerProSerPro 944  
 QY 851 --GGAACCTGCCAGGAGGAGGAGCAGCCCGAGTGAGGAGAGTACAGCTGAGTGTGA 794  
 Db 944 lYthrThrThrProGlyHisThrThraLathrSerAlaThrThrAlaThrProS 964  
 QY 793 CAGCCACAGAGGTTGGAGAGCAGTCATACCCCATGATTCTGTGACTAGACCCAGCAG 734  
 Db 964 erLysThrAlaThrSerThrLeu-----ProSerG 975  
 QY 733 TCCCATAGCTAGGCTGTAACTCACATTTCATTTGACATTATTCACGTTGTCTCT 674  
 Db 975 lInProThrSerAlaProIleThrThrThraValaThrMetGlyCysGluProGlnCysAla 995  
 QY 673 GGTACATCTGTCTGTCTG--GATGTAGCTTTGCTGTGAGGGT----- 630  
 Db 995 rPserGlu-----TriLeuAspTyrSerTyrProMetProGlyProSerGlyAla 1012  
 QY 629 -----TCCAGGTACATGACGTGCTGCTGAGTACAGTAC-----T 593  
 Db 1012 spHeAspThrTyrSerAsnIleArgAlaIleGlyAlaValaCysGluGlnProLeu 1032  
 QY 592 CCCTCCACAGCAGCAGCCAGCTGTGATAGTACAGTACAGTACAGCCGAGCCCTCAGATA 533  
 Db 1032 lYleuGlnCysArgAlaGln-Ala-----GlnProGlyValProLeuArgGluLeu 1048  
 QY 532 GGTACG----- 527  
 Db 1049 GlyGlnValaValaGlySerLeuAspPheGlyLeuValaCysArgAsnArgGluGlnVal 1068  
 QY 526 -----CCAGAGCCTTCACATATGTGCTTTG----- 497  
 Db 1069 GlyLysPheLeuMetCysPheAsnTyrGlnIleArgValaPheCysAsnTyrGlyHis 1088  
 QY 496 -----GATCAGGAGCACT 482  
 Db 1089 CysProSerThrProAlaThrSerSerThrAlaThrProSerSerThrProGlyThrThr 1108  
 QY 481 TGGCTCATCAGCTTGAAGAAAGACAGCTGCCAGCTGGGGGAGAGTACAGAGTA 422  
 Db 1109 TrpIleLeuThrGluLeuThrThrAla----- 1118  
 QY 421 ACAGGCTGATGTTCTCTTGTGAGCAGACACATTCCTCCAT 374  
 Db 1119 -----ThrThrThrGlySerThrAlaThrProThrSerThrLeuArg 1135  
 QY 373 -----TGAAGCTTGAAGACAGTATACCTCTCTCATAC 335  
 Db 1136 ThrAlaProProProLysValaLeuThrThrAlaThrThrProThrValThrSer 1155  
 QY 334 CTGAGCAGCCCTGTGATCGATGGGGATCATATCACTCCAGAGAGGTGC 275  
 Db 1156 LysAlaThrProSer-SerSerProGlyThrAlaThrAlaLeu-----Pr 1170  
 QY 274 TGGCATCTCCCTTGAAGGCTTCTCCAGTGC-----TCACTGTCCAGGTGT 224  
 Db 1170 AlaLeuArgSerThrAlaThrThrProThrAlaThrSerValaThrProIleProSer 1190  
 QY 223 CTGTTGTGACCAACAAGGCCGACAGTGGGCTTCACAGACATCCACAGCTC 164  
 Db 1190 rSerLeuGlyThrThr-----TprThrArgLeuSerGlnThrThrProTh 1206

QY 163 AGCCAGTTGACACT 148  
 Db 1206 rAlaThrMetSerThr 1211

## RESULT 8

EDBE11  
 immediate-early protein IE110 - human herpesvirus 1 (strain 17)  
 C:Species: human herpesvirus 1  
 C:Date: 31-Mar-1988 #sequence\_revision 31-Mar-1988 #text\_change 17-Mar-2000  
 C:Accession: A29152  
 J:Berry, L.J.; Rixon, F.J.; Everett, R.D.; Frame, M.C.; McGeoch, D.J.  
 J:Gen. Virol. 67, 2365-2380, 1986  
 A>Title: Characterization of the IE110 gene of herpes simplex virus type 1.  
 A:Reference number: A29152; NUID:87059760; PMID:3023529  
 A:Accession: A29152  
 A:Molecule type: DNA  
 A:Residues: 1-775 <PER>  
 A:Cross-references: GB:X04614; NID:G59832; PIDN:CAA28285.1; PID:G59833  
 C:Genetics:  
 A:Introns: 19/3; 242/1  
 C:Superfamily: herpesvirus immediate-early protein IE110; RING finger homology  
 C:Keywords: DNA binding; early protein; transcription regulation; zinc finger  
 F:112-162/Domain: RING finger homology <RNG>  
 F:116-156/Region: zinc finger C3HC4 motif

## Alignment Scores:

pred. No.:	0 209	Length:	775
Score:	117.00	Matches:	47
Percent Similarity:	36.00%	Conservative:	7
Best Local Similarity:	31.33%	Mismatches:	61
Query Match:	4.57%	Indels:	35
DB:	1	Gaps:	8

US-09-989-919-15 (1-1397) x EDBE11 (1-775)

QY 1188 CCGGCCCCCAGCCCGCCAGCTGCTGCTCAACACACCTGCTCCAGCCAGTC 1129  
 Db 381 ProGlySerAlaProArgProGlyProProAlaSerAlaAlaSer---GlyProAla 399  
 QY 1128 TTGGCTCATGCTGAGTG-----TGCACCTTGCCCTGACCCCTGGGCTGCTGGGA 1075  
 Db 400 ArgProArgAlaAlaValaAlaProCysValaArgAlaProPro----- 414  
 QY 1074 GTGTCCCTCTGTGGAGGTCAGAGTACGCTCCCGAGTACAGAAATCCACATCTCG 1015  
 Db 415 -----GlyProGlyProArgAlaProAla-ProS 424  
 QY 1014 AGCATCCCCCGCAAGTCTCTGTCAGACCTAAGCTGAGAGAACCTCCAGATGAGTT 955  
 Db 424 yAlaGluProAlaAlaArgProAlaAsp-----AlaArgVal----- 437  
 QY 954 GTCCCGACGAGCCCCCAGCCGCTGTGGCATGTGACAGCAGCAGCAGGAGGCTTGA 895  
 Db 438 -----ProGlnSer---HisSerSerLeuAlaGlnAlaAsnGlnGlnSerLeu 455  
 QY 894 GCTACATTCTCACATATTGAACACAGTATGG-----CTGCCACAGG 850  
 Db 455 sArgAlaArgAlaThrValaAlaArgGlySerGlyProGlyValaGlyGlyHis 475  
 QY 849 AACCGTCCCGAGGAGGAGCAGCCCGAGTGGGAGACTAGCTGAGCTTGCTTACAG 790  
 Db 475 y-----ProSerArgGlyAlaAlaProSerGlyAlaAlaProLeuProSerAlaAla 493  
 QY 789 CCACAGGTTGAGAGGACGATCAATCC 762  
 Db 493 rValaGlnGlnAlaAlaValaArgPro 502

## RESULT 9

atrophin-1 - human  
 G01763  
 C:Species: Homo sapiens (man)  
 C:Date: 21-Dec-1996 #sequence\_revision 06-Jun-1997 #text\_change 17-Jul-1998

C/Accession: G01763  
 R/Margolis, R.L.  
 Submitted to the EMBL Data Library, March 1995  
 A/Reference number: G08343  
 A/Accession: G01763  
 A/Status: preliminary; translated from GB/EMBL/DBJ  
 A/Molecule type: mRNA  
 A/Residues: 1-1184 <MAR>  
 A/Cross-references: EMBL:U23851; NID:915325; PID:915326  
 C/Genetics:  
 A/Gene: GDB:DRPLA; B37  
 A/Cross-references: GDB:270336; OMIM:125370  
 A/Map position: 12p-12p

## Alignment Scores:

Pred. No.:	0.209	Length:	1184
Score:	117.00	Matches:	92
Percent Similarity:	32.23%	Conservative:	34
Best Local Similarity:	23.53%	Mismatches:	140
Query Match:	4.57%	Indels:	125
DB:	2	Gaps:	20

US-09-989-919-15 (1-1397) x G01763 (1-1184)

```

QY 1190 GCCCGGCCCCCACC----- 1176
DB 365 AlaProAlaProProMetArgPheProTyrSerSerSerSerSerAlaAlaAla 384
QY 1175 -----CCGCCCAAGTCTGGCTCAAAACCA 1149
DB 385 SerSerSerSerSerSerSerSerSerSerAlaSerProPheProAlaSerGlnAlaLeu 404
QY 1148 CTTGCT-----CCCTGACCCAGCTTGGCTCATGCTGAGTGTGACCT 1104
DB 405 ProSerTyrProHisSerPheProProProProHisSerLeuSerValSerAsnGlnProPro 424
QY 1103 CTGCCCCCTGACCCCTGGGCTGGCTGGAGTGTCCCTGGGGAGGTGACAGATAGCCT 1044
DB 425 LysTyrThrGlnProSerLeuProSerGlnAlaValTrpSerGlnGlnProProProPro 444
QY 1043 CCCCAG-----GTACAGATACCCACATCTTGAGCATCCCGCCAG 999
DB 445 ProProTyrGlnAlaGlnLeuAlaAsnSerAsnAlaHisProGlyProPheProPro-- 463
QY 998 TTCCTGTTCAGACCTTACAGAGACACACAGAGAGGCTTATGACATG----- 939
DB 464 -----SerThrGlnAlaGlnSerThrAlaHisProProAlaSerThrHisHis 480
QY 938 CACCGCTGTGTGGCATGTGACAGACAGACAGAGGCTTATGACATG----- 888
DB 481 HisHisGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGln 500
QY 887 -----TTCACACTATTGAACACAGTATGGGCTGCC 855
DB 501 SerGlyProProProProGlyAlaPheProHisProLeuGlu-----GlyGlySerSer 518
QY 854 CACGGAACCTCCACAGGAGGAGGAGACAGCCCGCAGTGGAGACTAGCTGACCTTCTT 795
DB 519 HisHisAlaHisPro--TyrAlaMetSerPro-----SerLeuGlySerLeu 533
QY 794 ACAGCCACAGGGGTGGAGAGGAGCATATCCCATGATCTCTGACTAGAGACGACA 735
DB 534 ArgProTyrProProProGlyProAlaHisLeuProPro----- 545
QY 734 GTCCCATAGCTAGGCTGGTAACTCACAATTATTAACATTGGCATTTATGGACGTTTGC 675
DB 546 -----ProHis-----SerGlnValSer 551
QY 674 TGGTACATCTGTCTGTCTGGATGATTAAGTTTGTCTGACAGGGGTTCCACAGTGAATGC 615
DB 552 TyrSerGlnAlaGlyPro-----AsnGlyProProValSerSer 564
QY 614 AGTGTGCGACGTCAAGTCACTCCCTCACAGGACAGCCAGCTGCTGAGAGATAGTCAG 555

```

```

DB 565 SerSerAsnSerSerSer-SerThrSerGlnGlySerTyrProCys----- 579
QY 554 TGGTACCCGAGCCCTCAG-----ATAGTACGACGAGAGGCTTCATATATGTGT 504
DB 580 -SerHisProSerProSerGlnGlyProGlnGlyAlaProTyrProPheProProValPr 599
QY 503 CTTGTTGGATCAGGAGACCATTTGGTTCATCCAGCTTGAAAAAGACGAGCTCCGACC 444
DB 599 otherValThrThrSerSerAlaThrLeuSerThrValIleAlaThrValAlaSerSerPr 619
QY 443 TGGGGGCGACAGGTGACAGTAACAGGCTGATGTTCTCTTTCTGCA----- 395
DB 619 alaGlyTyrLys-----ThrAlaSerProProGlyProProProTyr 633
QY 394 -GGCAAGACCATTTCTCTTAATTGACTGCTTGAAGACAGTAACTCTCTTCATCA 336
DB 633 rGlyLysArg---AlaProSerProGlyAlaTyrLysThrAlaThrProProGlyTyrLys 652
QY 335 CCTGACACACCCCTCTGACTTCGAGTGGGGGATCACTATCTATCACTCCAGAGAGTTC 276
DB 652 sProGlySerProProSerPheArgThrGly-----ThrProPro----- 665
QY 275 CTGCCATCTCCCTTGAAGGCTTCTCCACATGCTACCTGTGCCAGGTGCTGTGG 216
DB 666 -----GlyTyrArgGly-----ThrSerProProAlaGlyProGln 677
QY 215 TGACACACAA-----GGCCGACACTGGGC 191
DB 677 TyrPheLysProGlySerProThrValGly 687

```

## RESULT 10

S50832  
 atrophin-1 - human  
 C/Species: Homo sapiens (man)  
 C/Date: 14-Jul-1995 #sequence\_revision 21-Jul-1995 #text\_change 07-May-1999  
 A/Accession: S50832  
 R/Nagafuchi, S.; Yanagisawa, H.; Ohaki, E.; Shitayama, T.; Tadokoro, K.; Inoue, T.; Yan  
 Nature Genet. 8, 177-181, 1994  
 A/Title: Structure and expression of the gene responsible for the triplet repeat disord  
 A/Reference number: S50832; PMID:7842016  
 A/Accession: S50832  
 A/Status: preliminary  
 A/Molecule type: mRNA  
 A/Residues: 1-1184 <NAG>  
 A/Cross-references: EMBL:D31840  
 C/Genetics:  
 A/Gene: GDB:DRPLA; B37  
 A/Cross-references: GDB:270336; OMIM:125370  
 A/Map position: 12p13.31-12p13.3112p-12p

## Alignment Scores:

Pred. No.:	0.209	Length:	1184
Score:	117.00	Matches:	92
Percent Similarity:	32.23%	Conservative:	34
Best Local Similarity:	23.53%	Mismatches:	140
Query Match:	4.57%	Indels:	125
DB:	2	Gaps:	20

US-09-989-919-15 (1-1397) x S50832 (1-1184)

```

QY 1190 GCCCGGCCCCCACC----- 1176
DB 366 AlaProAlaProProMetArgPheProTyrSerSerSerSerSerAlaAlaAla 385
QY 1175 -----CCGCCCAAGTCTGGCTCAAAACCA 1149
DB 386 SerSerSerSerSerSerSerSerSerSerAlaSerProPheProAlaSerGlnAlaLeu 405
QY 1148 CTTGCT-----CCCTGACCCAGCTTGGCTCATGCTGAGTGTGACCT 1104
DB 406 ProSerTyrProHisSerPheProProProProHisSerLeuSerValSerAsnGlnProPro 425

```





Db 370 SerAsnValIysPheValGlnAspThrSerLys-----PheTrp 382  
 QY 495 GATCAGGAGCACTTGCTGCATCC-----AGTCTGAAAAGACCACTGCCGAC 445  
 Db 383 -TylsPProHisLeuSerArgAspGlnAlaIleAlaLeuLysAspPro-- 401  
 QY 444 CTGGGGGGGCGAGGTAGAGGTAAACAGGGCTGATGTTCTTTCTGAGGCAAGACCA 385  
 Db 402 -----GlyAlaPheLeuIle--ArgAspSerH 410  
 QY 384 TTCTCCCTAATTGACTGCTTGAAAGACAGGTATACCTCTCTCATCACCAGACCC 325  
 Db 410 sSer-----PheGlnGlyAlaTyr-GlyLeuAlaLeuLysValAlaThrP 425  
 QY 324 CCTCTGACTTCCAGTGGGGGATCATCTATCATCACTCCAGAGAGTCTGCCATCTCC 265  
 Db 425 ro-----ProPheSerAlaGln-----P 431  
 QY 264 CTTGGAAGGCGC-----TTCTCCCACTGGCTC-----ACCTGTCGAC 229  
 Db 431 roTlrylsglyAspProValGlnGlnLeuValArgHisPheLeuIleGlnThrGlyProl 451  
 QY 228 GTGGT-----CTGTTGTGTGACCAACAGCCCGACACATGGGC-----ATGGCTCTCA 181  
 Db 451 ysglyValIysIleLysGlyCysProSerGlnProTyrPheGlySerLeuSerAlaLeuV 471  
 QY 180 CAGACATCCACAGCTCCTGAGCCAGTT 155  
 Db 471 alserGlnHisSerIleSerProIle 479  
 RESULT 12  
 A33380  
 Interleukin-4 receptor precursor - mouse  
 N/Alternate names: IL-4 receptor  
 C/Species: Mus musculus (house mouse)  
 C/Date: 30-Jun-1993 #sequence revision 30-Jun-1993 #text\_change 01-Dec-2000  
 C/Accession: A33380; B33380; G33380; A34861; I54232  
 R/Mosley, B.; Beckmann, M.P.; March, C.U.; Idzerda, R.L.; Gimpel, S.D.; Vandenbos, T.; H  
 Wimer, M.B.; Cosman, D.; Park, L.S.  
 Cell 59, 335-348, 1989  
 A/Title: The murine interleukin-4 receptor: molecular cloning and characterization of se  
 A/Reference number: A90911; MUID:90030408; PMID:2805066  
 A/Accession: A33380  
 A/Molecule type: mRNA  
 A/Residues: 1810 <MO1>  
 A/Cross-references: GB:W27959; NID:G198363; PIDN:AAA39299.1; PID:G309408  
 A/Accession: B33380  
 A/Molecule type: mRNA  
 A/Residues: 1-258 <MO2>  
 A/Accession: C33380  
 A/Molecule type: mRNA  
 A/Residues: 1224, PSNENL, <MO3>  
 A/Cross-references: GB:W27960; NID:G198365; PIDN:AAA39300.1; PID:G309409  
 A/Note: part of this sequence, including the amino end of the mature protein, was confir  
 A/Note: three forms of cDNA were isolated; the longest encodes extracellular, transmembr  
 acellular domain and may encode a soluble form of the receptor  
 R/Harada, N.; Castle, B.E.; Gotman, D.M.; Itoh, N.; Schreurs, J.; Barrett, R.L.; Howard,  
 Proc. Natl. Acad. Sci. U.S.A. 87, 857-861, 1990  
 A/Title: Expression cloning of a cDNA encoding the murine interleukin 4 receptor based c  
 A/Reference number: A34861; MUID:90138976; PMID:2405398  
 A/Accession: A34861  
 A/Molecule type: mRNA  
 A/Residues: 1810 <HAR>  
 A/Cross-references: GB:W29654; NID:G198346; PIDN:AAA39297.1; PID:G309407  
 R/Wrighton, N.C.; Campbell, L.A.; Lee, F.D.  
 Growth Factors 6, 103-118, 1992  
 A/Title: The murine interleukin-4 receptor gene: Genomic structure, expression and poten  
 A/Reference number: I54232; MUID:92265335; PMID:1534014  
 A/Accession: I54232  
 A/Status: Preliminary; translated from GB/EMBL/DBD  
 A/Molecule type: DNA  
 A/Residues: 1-73, 'I', '75-333, 'P', '335-810 <RES>  
 A/Cross-references: GB:W64879; NID:G198359; PIDN:AA859727.1; PID:G198361

C/Genetics:  
 A/Intons: 24/1; 71/2; 122/1; 172/3; 225/1; 258/2; 284/3; 301/2  
 C/Superfamily: Interleukin-4 receptor; cytokine receptor homology  
 C/Keywords: alternative splicing; cytokine receptor; glycoprotein; transmembrane protein  
 F/1-25/Domain: signal sequence #status predicted <SIS>  
 F/26-810/Product: interleukin-4 receptor #status experimental <Mat>  
 F/26-233/Domain: extracellular #status predicted <EX>  
 F/24-219/Domain: cytokine receptor homology <CRS>  
 F/234-257/Domain: transmembrane #status predicted <TM>  
 F/258-810/Domain: intracellular #status predicted <INT>  
 F/12,129,135,163,177/Binding site: carbohydrate (Asn) (covalent) #status predicted  
 Alignment Scores:  
 Pred. No.: 0.366 Length: 810  
 Score: 114.00 Matches: 102  
 Percent Similarity: 32.71% Conservative: 38  
 Best Local Similarity: 23.83% Mismatches: 123  
 Query Match: 4.45% Indels: 166  
 DB: 1 Gaps: 26  
 US-09-989-919-15 (1-1397) X A33380 (1-810)  
 QY 1133 CAGCTTGGCTGATGTGAGTGTGCACCTTGCCCTGACCCCTGG-----CTG 1083  
 Db 428 GlnSerAlaLeuAlaGlnSerCysSerProLeuProSerGlySerGlyGlnAlaSerVal 447  
 QY 1082 GCCTGGAGTGT-----CCCTGTGGGAGGTGACAGATGACCTCCCGATGACAGATCAG 1026  
 Db 448 SerTlryAlaCysLeuProMetGlyProSerGlnGlnAlaThrCysGlnValThrGlnGln 467  
 QY 1025 CCA-----CATCTGGA----- 1014  
 Db 468 ProSerHisProGlyProLeuSerGlySerProAlaGlnSerAlaProThrLeuAlaCys 487  
 QY 1013 ----- 1011  
 Db 488 ThrGlnValProLeuValLeuAlaAspAsnProAlaTyrArgSerPheSerAspCysCys 507  
 QY 1010 TCCCGCCCAAGTCTCTGTCCAGACCTTAGTGAGACACTCCACGATGATTG--- 954  
 Db 508 SerProAlaProAsnProGlyGlnLeuAlaProGlnGlnGlnAlaAlaPheHisLeuGln 527  
 QY 953 -----TCCCGACGAGCCCGCCCGCTG-----CTGGGCACTGG 918  
 Db 528 GlnGlnGlnProProSerProAlaAspProHisSerSerGlyProProMetGlnProVal 547  
 QY 917 CAGCAGCAGCAGCAGGCGCTTAGCTTACATCTTCACATTTGAAACACAGTATGGGCT 858  
 Db 548 GlnSerTlryGlnGln-----IleLeuHisMetSerValLeuGlnHis-----GlyAla 563  
 QY 857 GCCCAGGACCGTCCCGAGG----- 837  
 Db 564 AlaAlaGlySerThrProAlaProAlaGlyTyrGlnGlnPheValGlnAlaValLys 583  
 QY 836 -----AGGGCCACAGCCCGCAGTGGGAGACTAGCTG 804  
 Db 584 GlnGlyAlaAlaGlnAspProGlyValProGlyValAlaGlnProSerGlyAspProGlyTyr 603  
 QY 803 -----AGCTTGCTTACAGCCACAGGTT----- 780  
 Db 604 LysAlaPheSerSerLeuSerSerLeuSerSerLeuGlyIleArgGlyAspThrAlaAlaGly 623  
 QY 779 -----GGAGAGGACATCATCCCATATCTCTGACTAGAGCA----- 738  
 Db 624 ThrAspAspGlyHisGlyGlyTyrLysPro-----PheGlnAspProValProAsnGlnSer 642  
 QY 737 -----GCAGTCCCATAGCTAGCTGCTGTAACCTACATTTGACATTTGGCATTTATGCA 684  
 Db 643 ProSerSerValPro-----LeuPheThrPheGlyLeuAspThr 655  
 QY 683 CGTTGCTCTGCTGATCATCTGTC-----TGTCTGGA 652  
 Db 656 GlnLeuSerProSerProLeuAsnSerAspProProLysSerProProGlnCysLeuGly 675





QY 544 -AGCCCTCAGATGATGTCAGCCAGAGCCTTCACATANGTGTCTTGTGGATCAGGAC 486  
 Db 274 pATGATLeuArgpHeG1yAlaProGlnAlaPheLeuAlaGlyLeuLeuLeuAlaThrValAl 294  
 QY 485 CACTTGGCTTCATCCAGTCTTGAAGAACACCACTGCCAGCTGGGGGAGAGGTGAGA 426  
 Db 294 a-ValGlyThrAlaArgAlaGlyLeuGlnProArgAlaAspMet----- 308  
 QY 425 GGTAAACAGGGCTGGATGTTTCTCTTCTTGAGGCAAGACATTCCTTAATTGACTGC 366  
 Db 308 ----- 308  
 QY 365 CTGAAGACAGATATACCTCTCTCTATCACTCAGACCCCTTGACTTCGATGGG 306  
 Db 309 -----AlaAlaProProThr-LeuProGlnPro 317  
 QY 305 GATCACTATCTATCAACTCCAGAGAGTCCCTGCACTCCCTTGAAGGGCTTCCCA 246  
 Db 318 -----Pro 318  
 QY 245 CTGGCTCAGCTGTCAGAGTGTCTGTTGTTACACAAAGCCCACTGGGATGCG 186  
 Db 319 ArgAlaHisGlyGlnHisThrGlyHisHisHisGlnLeuProPheLeuGlyHisAsp 338  
 QY 185 TCTCAGACAGATCCAGAGCCTCAGCCAGGTTGAACACTGAAGAGGAGCAGCTCCCGT 126  
 Db 339 G1yHisHis-----GlyGlyHisLeuArgVal 347  
 QY 125 GGTAGATGTCAGAGCGGTAGT-----CTTCTGGGGAGTGG 87  
 Db 348 G1yGlnHisThrArgAlaAspValLeuProGlyHisThrLeuGlnGlyGlyTyr 367  
 QY 86 TGCTGTCTGGGA-----TACACTGTACTGGTACTGCG 54  
 Db 368 G1yCysThrAsnLeuSerAspTrpHisGlnGlyThrHisValCys 382

## RESULT 14

T09059  
 notch4 - mouse  
 C/Species: Mus musculus (house mouse)  
 C/Date: 11-Jun-1999 #sequence\_revision 11-Jun-1999 #text\_change 11-Jan-2000  
 C/Accession: T09059  
 R/Rowen, L.; Mahlreis, G.; Qin, S.; Ahearn, M.E.; Dankers, C.; Loretz, C.; Sc  
 submitted to the EMBL Data Library, October 1997  
 A/Description: Sequence of the mouse major histocompatibility locus class III region.  
 A/Reference number: Z16543  
 A/Accession: T09059  
 A/Status: preliminary; translated from GB/EMBL/DBJ  
 A/Molecule type: DNA  
 A/Residues: 1-1964 <ROW>  
 A/Cross-references: EMBL:AF030001; NID:G2564945; PID:G2564947  
 A/Genetics:  
 A/Gene: notch4  
 A/Map position: 17  
 A/Intons: 22/1; 49/2; 148/1; 264/1; 305/1; 384/1; 436/1; 501/1; 539/1; 577/1; 618/1; 67  
 1679/3; 1129/1; 1761/3  
 C/Superfamily: unassigned EGF-related proteins; EGF homology  
 C/Keywords: receptor; signal transduction  
 F/514-545/Domain: EGF homology <EGF>

## Alignment Scores:

Prod. No.: 0.365 Length: 1964  
 Score: 114.00 Matches: 128  
 Percent Similarity: 28.84% Conservative: 39  
 Best Local Similarity: 22.11% Mismatches: 191  
 Query Match: 4.51% Indels: 222  
 DB: 2 Gaps: 37

US-09-989-919-15 (1-1397) X-T09059 (1-1964)

QY 6 TGCACCTGTATCCGAGCGGCGAGATATCTGCAAGAACTCCAGCG-----CAAGACAGAGTA 59  
 Db 6 TGCACCTGTATCCGAGCGGCGAGATATCTGCAAGAACTCCAGCG-----CAAGACAGAGTA 59

Db 903 CysIleAspThrGlySerSerThrPheCysArgCysProProGlyPheGlnGlyLeu 922  
 QY 60 CCGAGTACAGAGTATCCAGACAGACAGACCCCGAG----- 97  
 Db 923 CysGlnAspAsnValAsnProCysGlnProAsnProCysHisHisGlySerThrCysVal 942  
 QY 98 -----GAAAGTAC-----CGTTGCTGGCCATCTTACACACCGAGAGTCCCTC 142  
 Db 943 ProGlnProSerGlyTyrValCysGlnCysAlaProGlyTyrGlnGlyGlnAsnCys 961  
 QY 143 CTTTCACTGTTAACTGGCTGAGCTGGATGTTCTGTGAAGCCATGCCAGTGTGG 202  
 Db 962 -----SerIleValLeuAspAlaCysGlnSerGln---ProCysHis 974  
 QY 203 GCCTTTGTGTGACCAACAGACACCTTGAAGAGTGGAGCCAGTGGAGAA-----GCC 256  
 Db 975 -----AsnHisGlyThrCysThrSerArgProGlyGlyPheHisCysAla 989  
 QY 257 CTTCCAGAGGAGATGGCAGACCTCTCTGAGAGTTG-----ATAGATAGT 301  
 Db 990 CysProProGlyPheValGly-----LeuArgCysGlnGlyAspValAspGlyCysLeu 1007  
 QY 302 GATCCCCATCGAAGTCAAGAGGGGTGCT-----GAGGTATGAGAGAGAGTAT 352  
 Db 1008 AspArgProCysHisProSerGlyThrAlaAlaCysHisSerLeuAlaAsnAlaPheTyr 1027  
 QY 353 ---ACGTGCTTCAGAGC----- 367  
 Db 1028 CysGlnCysLeuProGlyHisThrGlyGlnArgCysGlnValGlnMetAspLeuCysGln 1047  
 QY 368 AGTCAAAATGAGGAGAGATGCTTGTGCTTCAGAGAGAGAAATCCAGACCT-----GT 421  
 Db 1048 SerGlnProCysSerAsnGlyGlySerCys-GluIleThrThrGlyProProProGlyLys 1067  
 QY 422 TACCTTCACTCTGCCCCCAGTGGGAG----- 452  
 Db 1067 ThrHisHis---CysProGlyGlyPheGlnGlyProThrCysSerHisLeuAlaLeuSe 1086  
 QY 453 ---CTGTCCTTTTCAAGACTGATGAGAGGCAAGTGGTCCCTGATCCCAAGACAC 508  
 Db 1086 rCysGlyIleHisHisCysHisAsnGlyGlyLeuCysLeuProSerPro----- 1102  
 QY 509 ATATGGAAGGCTCTGCTGACTG-----ATCTGAGGCTTCGGTGAACA-- 554  
 Db 1103 -----LysProGlySerProProLeuCysAlaCysLeuSerGlyPheGlyGlyProAs 1120  
 QY 555 -----GCTGACTATCTCAGCAGCTGGGCTTCTCT-----GTGAGGAGGT 595  
 Db 1120 rCysLeuThrProProAlaProProGlyCysGlyProProSerProCysLeuHisAsnG1 1140  
 QY 596 GACTTGCACTGGCAGAGCTGCATGTCACCTGGGAACCCCTGAGACAAAGCTAACTCC 655  
 Db 1140 yThrCysThrGluThrProGly-----LeuGlyAsnProGlyPheGlnCysThrCysPr 1158  
 QY 656 AAGACAGACAGATGTGACAGACAAAGCTGCAAT----- 689  
 Db 1158 oProAspSer-----ProGlyProArgCysGlnArgProGlyAlaSerGlyCysGlnG1 1176  
 QY 690 -----AATGCCAATGTTAAATGATGAGTTTACAGCTAGCT 727  
 Db 1176 yArgGlyGlyAspGlyThrCysAspAlaGlyCys-----SerGlyProGlyGlyAs 1193  
 QY 728 ATGGAGC----- 734  
 Db 1193 pThrAspGlyGlyAspCysSerLeuGlyValProAspProTyrGlyCysProProHis 1213  
 QY 735 -----TCTGAGCTCTGATCCAGGAATCATG----- 761  
 Db 1213 sSerGlnCysThrLeuLeuPheArgAspGlyArgCysHisProGlnCysAspSerGlnG1 1233  
 QY 762 -----GGGTATAGTGCCTTCCAACTGTG----- 788  
 Db 1233 uCysLeuPheAspGlyTyrAspCys---GluIleProLeuThrCysIleProAlaTyrAs 1252





GenCore version 5.1.6  
Copyright (c) 1993 - 2003 Compugen Ltd.

OM nucleic - protein search, using frame\_plus\_n2p model

Run on: June 24, 2003, 13:47:02 ; Search time 119 Seconds  
(without alignments)  
4837.778 Million cell updates/sec

Title: US-09-989-919-15  
Perfect score: 2527  
Sequence: 1 ggtgcgcacccgtaccgga.....aaaaaaaaaagcggtc 1397

Scoring table: BLOSUM62  
Xgapop 10.0, Xgapext 0.5  
Ygapop 10.0, Ygapext 0.5  
Fgapop 6.0, Fgapext 7.0  
Delop 6.0, Delext 7.0

Searched: 671580 segs, 206047115 residues  
Total number of hits satisfying chosen parameters: 1343160

Minimum DB seq length: 0  
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Command line parameters:  
-MODEL=frame+ n2p model -DRV=xlp  
-Q/cgnt2\_1/USFTO.spool/US09989919/runat.24062003.102047.19184/app.query.fasta.1.1543  
-DB=SPREMBL\_21 -QFMT=fastan -SUFFIX=rspt -MINMATCH=0.1 -LOOPCL=0 -LOOPEXT=0  
-UNITS=bits -START=1 -END=-1 -MATRIX=blosum62 -TRANS=human40.cdi -LIST=45  
-DOCAUGN=200 -THR SCORE=pct -THR MAX=100 -THR MIN=0 -ALIGN=15 -MODE=LOCAL  
-OUTFMT=pcr -NORM=ext -HEAPSIZE=500 -MINLEN=0 -MAXLEN=200000000  
-USER=US09989919 @CGN 1.158 @runat.24062003.102047.19184 -NCPU=6 -ICPU=3  
-NO\_MAP -LANG=OTHER -NEG\_SCORES=0 -WAIT -DSPBL=LOCK=100 -LONGLOG  
-DEV\_TIME=OUT=120 -WARN\_TIME=OUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOP=6  
-FGAPEXT=7 -YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database :  
1: SPREMBL\_21:\*  
2: sp\_archaea:\*  
3: sp\_bacteria:\*  
4: sp\_fungi:\*  
5: sp\_human:\*  
6: sp\_invertebrate:\*  
7: sp\_mammal:\*  
8: sp\_mhc:\*  
9: sp\_organelle:\*  
10: sp\_phage:\*  
11: sp\_plant:\*  
12: sp\_rodent:\*  
13: sp\_varus:\*  
14: sp\_vertebrate:\*  
15: sp\_unclassified:\*  
16: sp\_virus:\*  
17: sp\_bacterioplasmid:\*  
18: sp\_bacteriophage:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

Result No.	Score	Query Match	Length	DB ID	Description
1	519.5	20.6	207	4 Q96109	Q96109 homo sapien

2	431	17.1	268	11 Q8VBZ6	Q8VBZ6 mus musculus
C 3	143	5.6	1394	4 O00512	O00512 homo sapien
C 4	135.5	5.3	631	6 Q9N1P1	Q9N1P1 bos taurus
C 5	134	5.2	4123	4 Q75851	Q75851 homo sapien
C 6	127.5	5.0	4957	4 Q14687	Q14687 homo sapien
C 7	127.5	5.0	5262	4 Q14686	Q14686 homo sapien
C 8	124.5	4.9	818	6 Q9N1P0	Q9N1P0 bos taurus
C 9	124.5	4.9	1280	11 Q9EPX2	Q9EPX2 mus musculus
C 10	123.5	4.9	669	4 Q75441	Q75441 homo sapien
C 11	122.5	4.8	647	10 Q9ARH1	Q9ARH1 brassica na
C 12	121.5	4.7	1400	4 Q9UP57	Q9UP57 homo sapien
C 13	120.5	4.7	763	4 Q9UPX0	Q9UPX0 homo sapien
C 14	120.5	4.7	1236	4 Q15398	Q15398 homo sapien
C 15	120	4.7	815	5 Q9V9W8	Q9V9W8 drosophila
C 16	119	4.6	457	5 Q961D0	Q961D0 drosophila
C 17	119	4.6	749	5 Q9W384	Q9W384 drosophila
C 18	118	4.7	704	11 Q9CWS0	Q9CWS0 mus musculus
C 19	118	4.6	1063	12 Q40956	Q40956 rubella vir
C 20	118	4.6	1063	12 Q8VA12	Q8VA12 rubella vir
C 21	118	4.6	1063	12 Q8VA10	Q8VA10 rubella vir
C 22	118	4.6	3570	4 Q99552	Q99552 homo sapien
C 23	117.5	4.6	596	11 Q9CVP3	Q9CVP3 mus musculus
C 24	117.5	4.6	749	5 Q9B1W6	Q9B1W6 drosophila
C 25	117	4.6	995	4 Q9Y2W4	Q9Y2W4 homo sapien
C 26	117	4.6	1182	4 Q9Y495	Q9Y495 homo sapien
C 27	116.5	4.6	970	12 Q8QZ06	Q8QZ06 rana tigris
C 28	116.5	4.6	1283	5 Q8T5H0	Q8T5H0 anopheles g
C 29	116.5	4.6	1511	4 Q75412	Q75412 homo sapien
C 30	116.5	4.6	1566	4 Q9P2R6	Q9P2R6 homo sapien
C 31	116	4.5	726	11 Q924S6	Q924S6 mus musculus
C 32	116	4.5	957	4 Q9UKN0	Q9UKN0 homo sapien
C 33	115.5	4.5	840	4 Q9Y2K5	Q9Y2K5 homo sapien
C 34	115	4.5	649	12 Q9J210	Q9J210 macaca mla
C 35	115	4.5	699	4 Q9NT29	Q9NT29 homo sapien
C 36	115	4.5	907	12 Q6E537	Q6E537 human herpe
C 37	115	4.5	961	3 Q92223	Q92223 emericella
C 38	115	4.5	1012	4 Q43393	Q43393 homo sapien
C 39	115	4.5	1012	4 Q73539	Q73539 homo sapien
C 40	115	4.5	1400	11 Q9ESU6	Q9ESU6 mus musculus
C 41	115	4.5	1400	11 Q9VHP8	Q9VHP8 mus musculus
C 42	114.5	4.5	1063	12 Q991E6	Q991E6 rubella vir
C 43	114.5	4.5	1063	12 Q991E4	Q991E4 rubella vir
C 44	114	4.5	585	12 Q41935	Q41935 murid herpe
C 45	114	4.5	810	11 Q60583	Q60583 mus musculus

ALIGNMENTS

RESULT 1	PRELIMINARY:	PRT:	207 AA.
Q96109	Q96109		
AC	Q96109		
DT	01-DEC-2001 (TREMBLrel. 19, Created)		
DT	01-DEC-2001 (TREMBLrel. 19, last sequence update)		
DT	01-DEC-2001 (TREMBLrel. 19, last annotation update)		
DE	Unknown (Protein for IMAGE:4139786) (Fragment).		
OS	Homo sapiens (Human).		
OC	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;		
OC	Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.		
OX	NCBI_TaxID=9606;		
RN	(1)		
RP	SEQUENCE FROM N.A.		
RC	TISSUE=MUSCLE;		
RA	Straussberg R.;		
RL	Submitted (MAY-2001) to the EMBL/GenBank/DBJ databases.		
DR	EMBL; BC007901; AAH07901.1; -		
FT	NON TER		
FT	1		
FT	1		
SO	SEQUENCE	207 AA;	23189 MW; 74335A78000B16B0 CRC64;
Alignment Scores:			
Pred. No.:	3.18e-39	Length:	207
Score:	519.50	Matches:	105
Percent Similarity:	59.66%	Conservative:	0

Best Local Similarity: 59.66% Mismatches: 1  
 Query Match: 20.56% Indels: 70  
 DB: 4 Gaps: 1

US-09-989-919-15 (1-1397) x 096109 (1-207)

QY 2 GTGCTCACCTGATACCGGAGGGGAGTATCTGCAGAACTCCAGGCAAGCAGCAGTACC 61  
 DB 102 ValLeuHisLeuYrnsrserglyGlnYrLeuGlnAnserThrAlaSerSerThr 121  
 QY 62 GAGTACCACTGATATCCAGACAGACCATCCCGAGAGAGACTACCGCTGTGGCATCC 121  
 DB 122 GluTyGlnCysrIleProAspSerThrIleProGlnGlnAspYrnsrCysrIleProSer 141  
 QY 122 TACCAACCAAGGAGGCTGCTCTTTCAGGTTCACCTGGCTGAGGCTGTGATGTCTGT 181  
 DB 142 TyrHisHisIleGlySerCysLeuLeuSerValPheAsnLeuAlaGlnAlaValAspValCys 161  
 QY 182 GAGAGCATGCGCCAGTGTGGGCTTGTGTGCACCAACAGACCACTGCAGAGTGA 241  
 DB 162 GluSerHisAlaGlnCysrAlaPheValAlaThrAsnGlnThrTrpThrIlePyrGlyArg 181  
 QY 242 CCAGTGGAGAAAGCCCTTCCAGAGAGATGGCAGAGACCTCTGTGAGTTGATAGTAGT 301  
 DB 182 -Gln----- 182  
 QY 302 GATCCCCCATCGAAGTCAGAGGGGGTGTGAGTATGAGAGAGATACGTGTCTT 361  
 DB 182 ----- 182  
 QY 362 CAGGCACTCAATTAAGGAGAAATGGTCTTGCTCCAGAAAGAAACATCCAGCCCTGT 421  
 DB 182 ----- 182  
 QY 422 TACCTCTCACTCTGTGCCCCCAGAGTGGCAGCTGTCTTTTCAAGACTGATGAGACCA 481  
 DB 183 -----LeuValPhePheIleThrIleGlyTrpSerGly 192  
 QY 482 AGTGTCTCTGATTCACCAAGACCATATGTGAAGGCTCTTGGC 527  
 DB 192 nValIvalProAspProAsnLysThrThrIleValIleValSerGly 207

## RESULT 2

QY 08VBZ6 PRELIMINARY; PRT; 268 AA;  
 AC 08VBZ6;  
 DT 01-MAR-2002 (TReMBLrel) 20, Created  
 DT 01-MAR-2002 (TReMBLrel) 20, Last sequence update  
 DT 01-JUN-2002 (TReMBLrel) 21, Last annotation update  
 DE Hypoetical 30.1 kDa protein (Fragment).  
 GN A115348.  
 OS Mus musculus (Mouse).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
 OX NCBI\_TaxID=10090;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC TISSUE=LIVER;  
 RA Strassberg R.;  
 RL Submitted (JAN-2002) to the EMBL/GenBank/DBJ databases.  
 DR EMBL: BC022157; AAR22157.1;  
 DR MGD: MGI:2146795; A115348.  
 KW Hypoetical protein.  
 FT NON TER 1  
 SQ SEQUENCE 268 AA; 30090 MW; 688B66E2C929E33C CRC64;

## Alignment Scores:

Pred. No.: 4.81e-31 Length: 268  
 Score: 431.00 Matches: 90  
 Percent Similarity: 53.71% Conservative: 4  
 Best Local Similarity: 51.43% Mismatches: 10  
 Query Match: 17.06% Indels: 71  
 DB: 11 Gaps: 2

US-09-989-919-15 (1-1397) x 08VBZ6 (1-268)

QY 5 CTGCACCTGTACCGGAGGGGAGATATCTGCAGAACTCCAGGCAAGCAGCAGTACCNG 64  
 DB 165 LeuHisLeuPheAsrserglyGlnYrLeuGlnAnserThr---SerSerGlnAlaGln 183  
 QY 65 TACCACTGATCCAGACAGACCATCCCGAGAGAGACTACCGCTGTGGCATCC 124  
 DB 184 TyrGlnAlaGlnIleProAspSerAlaIleThrGlnGlnAspYrnsrCysrIleProSer 203  
 QY 125 CACCAAGGAGAGCTGCTCTTTCAGGTTCACCTGGCTGAGGCTGTGATGTCTGTG 184  
 DB 204 HisHisGlyGlyCysLeuLeuSerValPheAsnLeuAlaGlnAlaIleAspValCysGln 223  
 QY 185 AGCCATGCCCATATGTGCGGCTTGTGTGCACCAACAGACCACTGCAGAGTGA 244  
 DB 224 SerHisAlaGlnCysrAlaPheValAlaThrAsnGlnThrTrpThrIlePyrGlyArg-Ly 243  
 QY 245 GTGGAGAAAGCCCTTCCAGAGAGATGGCAGAGACCTCTGTGAGTTGATAGTAGT 304  
 DB 243 -S----- 243  
 QY 305 CCCCATCGAAGTCAGAGGGGGTGTGAGTATGAGAGAGAGTATACGTCTTCAA 364  
 DB 243 ----- 243  
 QY 365 GGCAGTCAATTAAGGAGAAATGGTCTTGCTCCAGAAAGAAACATCCAGCCCTGTAC 424  
 DB 243 ----- 243  
 QY 425 CTGTCACTCTGCCCCCAGAGTGGCAGCTGTCTTTTCAAGACTGATGAGACCACT 484  
 DB 244 -----LeuValPhePheIleThrIleGlyTrpSerGlnAla 254  
 QY 485 GGTCTCTGATTCACCAAGACCATATGTGAAGGCTCTTGGC 527  
 DB 254 IvalIvalProAspAlaGlyLysThrThrIleValIleValSerGly 268

## RESULT 3

QY 000512 PRELIMINARY; PRT; 1394 AA;  
 AC 000512;  
 DT 01-JUL-1997 (TReMBLrel) 04, Created  
 DT 01-JUL-1997 (TReMBLrel) 04, Last sequence update  
 DT 01-DEC-2001 (TReMBLrel) 19, Last annotation update  
 DE B-cell CLL/Lymphoma 9.  
 GN BCL9.  
 OS Homo sapiens (Human).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Primates; Catarrhini; Hominiidae; Homo.  
 OX NCBI\_TaxID=9606;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RA Willis T.G., ZALCBERG I.R., Coignet L.J.A., Wlodarska I., STUL M.,  
 RA Jadavvel D.M., Bastard C., Treleaven D.G., Catovsky D., Silva M.L.M.,  
 RA Dyer M.J.S.;  
 RT "Molecular cloning of translocation t(11;14) (q21;q32) defines a novel  
 gene (BCL9) at chromosome 1q21.";  
 RL Blood 91:1871-1873 (1998).  
 DR EMBL: Y13620; CAA73942.1;  
 KW B-cell.  
 SQ SEQUENCE 1394 AA; 145970 MW; 20F75C121627E7BC CRC64;

## Alignment Scores:

Pred. No.: 0.000229 Length: 1394  
 Score: 143.00 Matches: 118  
 Percent Similarity: 31.03% Conservative: 39  
 Best Local Similarity: 23.32% Mismatches: 161  
 Query Match: 5.59% Indels: 188  
 DB: 4 Gaps: 24

US-09-989-919-15 (1-1397) x 000512 (1-1394)

```

QY 1221 GGGTTTCATGAAAGCAATGAGCAGAAAGCCCGCCGCCACCC----- 1174
DB 770 GYGLHHSproIngin-----GluTYGLymeGlyProArgProPheLeuPromet 787
QY 1173 GCCCCAGGTCCTGGCTCA----- 1156
DB 788 SerGInglyProGlySerAsnSerGlyLeuArgAsnLeuArgLubProIleGlyProAsp 807
QY 1155 -----AACCAACCTGCTCCTCGACCCCGAGTCTTG-----GGTCAT 1120
DB 808 GlnArgThrAsnSerArgLeuSerHisMetProProLeuAsnProSerSerAsn 827
QY 1119 GCTAGAGGTGACCTCTGCCCCCTGACCCCTGGGCTGGAGAGTGTCCCTGTGCG 1060
DB 828 ProThrSerLeuAsnThrAlaProProValGlnArgGlyLeuGlyArgLysProLeuAsp 847
QY 1059 AGGTCAAGATGACCTCCCGAGTACAGATCAACCATCTCGAGACATCCCGCCCAA 1000
DB 848 IleSerValAlaGlySer-GlnValHisSer-----ProGlyYlIeAsnProLeuLys 864
QY 999 GTCTCTGTCACAGCTAGAGTACAGAAACTCCAC----- 963
DB 864 SerPro-----ThrMetHisGlnValGlnSerProMetLeuG1 877
QY 962 -----GATGATTTGTCACAGCAGAGCCCGCTGCTGTCAGATGCA 916
DB 877 ySerProSerGlyAsnLeuLysSerProGlnThrProSerGlnLeu-AlaGlyMetLeuA 897
QY 915 GCAGCAGCAGCAGGGGCTTGTAGTACATTCTCACTATTGAAACAGATGATGGGCTGC 856
DB 897 laGlyProAlaAlaAlaAlaSerIle----- 905
QY 855 CCACGGAACCGTCCACAGAGGGGACAGCCCGCTGAGAGTACCTAGCTTGCT 796
DB 906 -----LysSerProProVal-----LeuGlySerAlaAla 916
QY 795 TACAGCCACAGGGTTGAGAGAGCAGTCAATACCCCATGATTC-----TGACTAGCA 742
DB 916 laSerProValHisLeuLysSerProSerLeuProAlaProSerProGlyTyrThrSers 936
QY 741 GCCAGAGTCCCATACCTAGCTGTGTAACCTCACTTTTAACATTGGCATTTAGCAGC 682
DB 936 exProLysProPro----- 940
QY 681 TTGTCTCTGTCATCTGTCTGTCTGGAGTTAGCTTTGTCTGACGGGTTCCAGGT 622
DB 941 -----LeuGlnSerProGlyYlIeProProA 949
QY 621 GACATCAGTGTCCAGTGCAGTGCATCTCCCTCC-----ACAGGCAAGCCAGCTGCTGAG 565
DB 949 snHisLysAlaProLeuThrMetAlaSerProAlaMetLeuGlyAsnValGln----- 966
QY 564 GATAGTCAAGTGTGACCGGAGCCCTCAGATAGTACAGCAGAGGCTTTCACATATGTG 505
DB 967 -----SerGlyLysProProProProThrAlaSerGlnProAlaSerValAsnIleProG 985
QY 504 TCTTTGTGGATGAGGACCACTTGCTCCATCAGTCTTGAAGAACAGAGCTGCCGAC 445
DB 985 LysLeuProSerSerThrProLysThrMetProProGluProThrIleSerGlnAsnP 1005
QY 444 CTGGGGGGGAGAGTGAAGAGTAAAGGCTGATGTTTCTCTTCTGCA-----GGCAAGA 388
DB 1005 ro-----LeuSerIleMetMetSerArgMetSerLysP 1016
QY 387 CCATTCCTCCTAATTGACTGCTTGAAGACAGTATACCTCTCTCATCAGCTCA----- 331
DB 1016 heAlaMetProSerSerThr-ProLeuTyrHisAspAlaIleLysThrValAlaSerSer 1035
QY 330 -----GCACCCCTCTGACTTCGATGG-----GGATCA 301
DB 1036 AspAspAspSerProProAlaArgSerProAsnLeuProSerMetAsnAsnMetProGly 1055

```

```

QY 300 CTATCTATCAACTCCAGAGAGGTCTGCCATCTCC----- 265
DB 1056 MetGlyYlIeAsnThrGln-----AsnProArgIleSerGlyProAsnProValAlaPromet 1074
QY 264 -----CTTGAGAGGCTTCTCCACAGTGCCTCACTTCCAGGTG----- 225
DB 1075 ProThrLeuSerProMetGlyMetThrGln-ProLeuSerHisSerAsnGlnMetProse 1094
QY 224 -----TCTGTTGTG 214
DB 1094 rProAlaAlaValGlyProAsnIleProProHisGlyValPrometGlyProGlyLeu 1114
QY 213 ACCACAAAGCCCGACACTGGGAGTGGCTTCACAGACATCCACAGCTTACAGCAGT 154
DB 1114 tSerHisAsnProIleMetGlyHisGlySerGlnGluProPrometValProGlnGly 1133
QY 153 AACACTGAAGAAGAGGAGCTCCCGTGT----- 126
DB 1134 -----ArgMetGlyPheProGlnGlyPheProProValGlnSerProProGlnGlnVa 1151
QY 125 -----GGTAGATGCGCAGCAGCGGTAGTCTTCTGGGGGATGTCGTC 82
DB 1151 lProPheProHisAsnGlyProSerGlyGlyGlnGlySerPheProGlyGly 1168
QY 81 TCTGGATACACTGTACT-----CGTACTCTGCTGCTGCGGAGTTCGTC 34
DB 1169 -MetGlyPheProGlyGlyGlyGlyProLeuGlyArgProSerAsnLeuProGlnSerSer 1188
QY 33 AGATCTGCC 24
DB 1188 aaSpAlaAla 1191

```

## RESULT 4

```

ID Q9N1P1 PRELIMINARY, PRT, 631 AA.
AC Q9N1P1;
DT 01-OCT-2000 (TrEMBLrel, 15, Created)
DT 01-OCT-2000 (TrEMBLrel, 15, Last sequence update)
DE 01-JUN-2002 (TrEMBLrel, 21, Last annotation update)
DE Submaxillary mucin (Fragment).
GN BSM1.
OS Bos taurus (Bovine).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovidae;
OC Bovidae; Bovinae; Bos.
OX NCBI_TaxID=9913;
RN (1)
RP SEQUENCE FROM N.A.
RC STRAIN=BRED ANGUS;
RX MEDLINE=20223253; PubMed=10759843;
RA Jiang W., Gupta D., Gallagher D., Davis S., Bhavanandan V.P.;
RT "The central domain of bovine submaxillary mucin consists of over 50
RT tandem repeats of 329 amino acids: chromosomal localization of the
RL BSM1 gene and relations to ovine and porcine counterparts.";
RL Eur. J. Biochem. 267:2208-2217(2000).
DR EMBL: AF178427; AAF67278.1; -
DR InterPro: IPR000194; ATPase_a/bcentre.
DR PRINTS: PRO1561; EDG8RECEPTOR.
DR PROSITE: PS00152; ATPASE_ALPHA_BETA; UNKNOWN_1.
FT NON TER 1
FT NON TER 1
SQ SEQUENCE 631 AA, 57067 MW, 184E9F0C2436115A CRC64;

```

## Alignment Scores:

```

Pred. No.: 0.000982 Length: 631
Score: 135.50 Matches: 123
Percent Similarity: 31.88% Conservative: 30
Best Local Similarity: 25.62% Mismatches: 147
Query Match: 5.29% Indels: 180
DB: 6 Gaps: 19

```

US-09-989-919-15 (1-1397) x Q9N1P1 (1-631)

QY 1191 GGGCCCCGCCCCCCCCCGAGGCTCTGCTCAACACCACTGCTC----- 1144  
 DB 50 GYGLYAlaValThrProAlaProSerProGlyGlySerSerAlaThrAlaGlyHisGly 69  
 QY 1143 -----TCCCTGACCCCGAGTCTGGCTCAT-----GCTAGAGTGTCACCTCT 1102  
 DB 70 ValGlySerAlaThrValAlaGlnAlaSerGlyAlaThrGlyAlaAspValLeuArgSer 89  
 QY 1101 GCGCCCTGAC-----CCCTGGGCGTGGCTGGAGTGTCCCTGTGGGAGGTGAGATAGCC 1045  
 DB 90 GlyThrSerLeuProValSerGlyValAlaValSerProGly-----Ser 104  
 QY 1044 TCCCGAGGTACAGAAATCACCAATCCTGGAGCATCCCGCCAGATCTCTGTCCAGAC 985  
 DB 105 SerProGlyArgSerGlyAlaThrAla----- 113  
 QY 984 CTAGCTGAGAGAACCTCCAGCATGATGTCCTCCAGCGAGCCCGCCTGTGTG 925  
 DB 114 -----ValSerSerGlnGlySerGlnProThrVal 123  
 QY 924 GCATGTGCAGACGACGAC-----GCAGGGCTTTAGTACATCTCTCA 883  
 DB 124 AlaLeuSerGlyAlaThrGlyThrSerValGlyProSerGlyThrArg-PheSerSerSe 143  
 QY 882 CACTATTGAACACAGATGATGGGTGCGCCAGGAGACCGTCCAGGAGGAGGACACA----- 828  
 DB 143 rAlaIleProAlaThrProGlySerThrThrGlyArgAlaAlaGlyAlaGlyThrProG 163  
 QY 828 ----- 828  
 DB 163 yValaSerGlnGlnSerAlaSerLeuAlaAlaAlaArgProThrAlaLeuGlyPyr 183  
 QY 827 -----GCCCCAGTGGGAGAGACTAGCTGAGCTGCTTACAGCCACAGGGTGG 778  
 DB 183 oglyThrSerAlaProSerGlyGlnThrSer-GlnSerArgSerSerValProGly-- 201  
 QY 777 AGAGGAGTATATACCCCATGATCTCTGACTAGAGAGCGACGATCCCATAGCTAGGCTG 718  
 DB 202 ---GlySerGlnThrGlnGlnProGlyThrGlySerGlnSerProThrLeuSerLeu 220  
 QY 717 ---GTAACCTACATTTTAACTTTGACATTGACATTGACGTTTCTGCTGACATCTGTC 661  
 DB 221 GlyValThrArgThrThrAlaLeuArgGlySerGlyThrArgValProSerThrGlyVal 240  
 QY 660 TGTCTGGAGTATAGCTTCTCTGACAGGGTTCACAGGTACATGACATGCTCCAGTGC 601  
 DB 241 -----SerGlyLeuProGlySerThrGln-GlyGlySerAl 252  
 QY 600 AAGTCACT----- 593  
 DB 252 aaAla-ThrGlyGlySerGlyAlaGlySerGlyProThrAlaProValSerGlyGlnThra 272  
 QY 592 -----CCCTCACAGGCAAGCCAGCTGTGAGAT 562  
 DB 272 rGThrIleValIleSerGlyThrAsnValProValSerGlyAlaProValThrProGlys 292  
 QY 561 AGTCACTGTGACGCG--AGCCCTCAATAGTACAGCCAGAGGCTTACATATATG 505  
 DB 292 eSerIleGlySerSerGlyAlaProGlyAlaGlyGlyProGlySerGlnThrAlaSerP 312  
 QY 504 TCTTGTGGGATAGGAGACACTGGCTGCATCCAGCTTGAAAGAACAGACGTCGCGAC 445  
 DB 312 roleuSerGlyAlaAlaGlyTh----- 319  
 QY 444 CTGGGGGGGAGAGGTAGAGGTAAAGGGCTGATGTTCTCTTTCTGGAGGAGACCA 385  
 DB 320 -----SerAlaThrGlySerGlyThri 327  
 QY 384 TTTCTCCCTAATTGACTGCTTGAAGACAGTATACTCTCTCATCACTACCTAGACACC 325  
 DB 327 leIleProProSerGlyAla-----Pro-ValThrProGlnPro 319  
 QY 324 CCTCTAGCTTCGATGGGAGTACATATATCAACCTCAGAGAGGTCTCTGCACTCTC 265

DB 340 ProIleIleSerThrGlyAlaSerAlaGlyProSerAlaSerSerGlnSerThrValThr 359  
 QY 264 CTGGAGAGGCTTTCTCCACTGCTTCACCTGCTTCACCTGCTGTTGATGACACCAAG 205  
 DB 360 LeuProGlyAla-----ThrGlyThrAspVal----- 368  
 QY 204 GCGCCGACACTGGGAGTGGCTTCACAGACATCCAGCTTCAGCCAGGTTGAACACTGAA 145  
 DB 369 -----LeuArgSerGlyThrSerLeu-ProValSerGlyGlyAlaValTh 383  
 QY 144 AGAGGACACTCCCTGCTGTGATGATGCGACAGACGCTGCTTCTGGGAGATGTTG 85  
 DB 383 rProAlaSerSerProGlyGlySerSerAlaThrAlaGlyProAlaValGlySerGlnTh 403  
 QY 84 CTGTCTG-----GATACACTGTACTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 34  
 DB 403 rThrValGlnValSerGlyAlaThrGlyThr----- 413  
 QY 33 AGATACCTGCGCGCTCGGT-----ACAGGTGACGCA 3  
 DB 414 -AspValLeuArgSerGlyAlaSerLeuProValSerGlyAlaAla 428  
 RESULT 5  
 ID 075851 PRELIMINARY; PRT; 4123 AA.  
 AC 075851;  
 DT 01-NOV-1998 (TREMBLrel. 08, Created)  
 DT 01-NOV-1998 (TREMBLrel. 08, Last sequence update)  
 DE 01-JUN-2002 (TREMBLrel. 21, Last annotation update)  
 GN MUGSC:H.DJ0751H13.1 protein (Fragment).  
 OS Homo sapiens (Human)  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Primates; Catarrhini; Hominiidae; Homo.  
 OX NCBI\_TaxID=9606;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RA Leonard S., Graves T., Stromatt C.;  
 RT "The sequence of Homo sapiens PAC clone RP4-751H13.";  
 RN Submitted (JUN-1998) to the EMBL/GenBank/DBJ databases.  
 RA SEQUENCE FROM N.A.  
 RA Waterston R.;  
 RL Submitted (DEC-1999) to the EMBL/GenBank/DBJ databases.  
 DR EMBL; AC004877; AAC36301.1; --  
 DR HSSP; P01130; 1AUJ.  
 DR InterPro; IPR000923; BlueCu.1.  
 DR InterPro; IPR001064; Crystallin.  
 DR InterPro; IPR000561; EGF-like.  
 DR InterPro; IPR000421; FAS8.C.  
 DR InterPro; IPR001092; HLH\_Basic.  
 DR InterPro; IPR002123; Kunitz\_Bpt1.  
 DR InterPro; IPR002172; LDL\_recept\_A.  
 DR InterPro; IPR002919; TIL\_Cyexich.  
 DR InterPro; IPR000884; TSP1.  
 DR InterPro; IPR001007; VWF\_C.  
 DR InterPro; IPR001846; VWF\_D.  
 DR Pfam; PF00754; F5\_P8\_type\_C.1.  
 DR Pfam; PF00057; Ild\_recept\_a.11.  
 DR Pfam; PF01826; TIL; 5.  
 DR Pfam; PF00090; TSP\_1; 14.  
 DR Pfam; PF00094; VWF\_3.  
 DR PRINTS; PR00261; LDRECEPTOR.  
 DR SMART; SM00231; FAS8C.1.  
 DR SMART; SM00192; LDla; 10.  
 DR SMART; SM00209; TSP1; 14.  
 DR SMART; SM00214; VWC; 1.  
 DR SMART; SM00216; VWD; 3.  
 DR PROSITE; PS00280; BPTI\_KUNITZ\_1; UNKNOWN\_1.  
 DR PROSITE; PS00196; COPPER\_BLUE; UNKNOWN\_1.  
 DR PROSITE; PS00225; CRYSTALLIN\_BETAGAMMA; UNKNOWN\_1.  
 DR PROSITE; PS00022; EGF\_1; UNKNOWN\_1.



DR PROSITE; PS01209; LDRA\_1; 9.  
DR PROSITE; PS0068; LDRA\_2; 18.  
DR PROSITE; PS0092; TSP1; 22.  
KW Glycoprotein.  
FT NON TER  
SQ SEQUENCE 1 1  
4123 AA, 434981 MW, 7AAB6FE9DC012FB CRC64;  
Alignment Scores:  
Pred. No.: 0.00187 Length: 4123  
Score: 134.00 Matches: 110  
Percent Similarity: 33.41% Conservative: 36  
Best Local Similarity: 25.17% Mismatches: 134  
Query Match: 5.23% Indels: 157  
DB: 4 Gaps: 26  
US-09-989-919-15 (1-1397) x 075851 (1-4123)  
QY 1249 GGGGAAAGCCTGTTGGCTGCTTGGAGCTTTGATGAAAGCAATGAGCGAAGG 1190  
DB 1799 GYGLVLAleuProserGlySerLeuValLeuSerLeu-----AspArg 1813  
QY 1189 CCCCCGCCCCGCGCCGAGCTGCTCAACACACCTGCTCCCTGACCCCACT 1130  
DB 1814 ProAlaAlaHisProProProProser-GlySerAspCysTrpProSerLeuSerGlyLe 1833  
QY 1129 CTTGGCTGATGCTAGAGGTGACCTCTGCCCTGACCCCTGGAGCTGCGCTG- 1076  
DB 1833 u-TrpLeuValLeu-----LeuValThrLeuGlyValProGlyProLeuTrpLeuProG 1852  
QY 1075 AGTGTCCC-----CTGTGGGAGGTGACAGATAGCTCCCGAGTACAGATACCCAC 1022  
DB 1852 LmHisProValValLeuProGluLeuGlnGlnProProLeuArgProArgSerProV 1872  
QY 1021 ATCTGAGAGATCCCGCCGAGCTCTCTGCTGACACCT-----ANG 980  
DB 1872 aLProTrp-HisProProGlnGlyTrpGlnThrGluProCysGluGlyCysGluHisGln 1891  
QY 979 CTGAGGAACACTCCAGATGATGATTCCTCCAGACGCCCGCCCTGCTGTCGACG 920  
DB 1892 GYGLVLAleuHisArgValGlyGluArgTrpHisGlyGlyPro-----CysArgVal 1908  
QY 919 TGCAGACGACGACGAGGGGCTTACATTTCTACATTTATTAACATGATGGG 860  
DB 1909 CysGln-CysLeuHisAsnLeuThrAlaHisCysSerProTy- 1922  
QY 859 CTGCCACCGAACCCTCCAGGAGGGGACACGCCCCAGTGGGAGACTAGCTTACCT 800  
DB 1923 -CysPro-----Le 1925  
QY 799 TCTTACAGCCACAGGTTGAGAGGACAGTATACCCCATGATTCCTGAGTACGAG 740  
DB 1925 uGlySerCysProGlnGlyTrp-----ValLeuValGluGlyThrGlyGly 1940  
QY 739 CAGCAGTCCCATGAGTGTGTAACATCAATTTAATTTGACATTTATTCAGCT 680  
DB 1940 user 1941  
QY 679 TGTCTGTGATCATCTGTCTGTCTGTGATTTAGCTTTGTCTGAGGGGTCCAGGTGA 620  
DB 1942 -----CysCysHisCys-----AlaLeuProGlyGly 1950  
QY 619 CATGACAGTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 569  
DB 1950 uAnGlnThrValGlnProMetAlaThrProAlaAlaAlaProAlaProSerProGln-1 1970  
QY 568 TGAAG-----ATAGTACAGTGTGACCGGAGCTTCAATAGTACGAGCGAGGCTTC 515  
DB 1970 LAlrGheProLeuAlaThrTyrlleuProProserGlyGlySerCysArgProLeu 1990  
QY 514 AC-----ATATGT-----G 506  
DB 1990 eSerProThrProAlaCysLeuSerLeuLeuHisProAspProCysTyrSerProLeu 2010

QY 505 GTCTTGTGGATCAGGAC-----CACTTGCTCCATCCAGCTTTGAAAAAGC 455  
DB 2010 LyeuAlaGlyLeuAlaGlnGlySerLeuHisAlaSerSerGlnGlnLeuGlnHisPro 2030  
QY 454 AG-----CTCCGACCTGGGGGACAGAGTGAAGGTAAAGGCTG 413  
DB 2030 hGlnAlaAlaLeuLeuGlnGlyAlaProThrGlnGlnProSerProGlnGlyTrpHisAla 2050  
QY 412 GA-----TGTTCCTCTTCTGAGG-----CAAGACATTCCTCCATTTAGTGCCTG 362  
DB 2050 YGlyAspAlaTyrlaAlaGlyTrpHisThrArgProHisTyrlleuGlnLeuAspLeuG 2070  
QY 361 AA-----GACACCTAATCT 347  
DB 2070 LmProArgAsnLeuThrGlylleuValProGluThrGlySerSerAsnAlaTyrlaAs 2090  
QY 346 CTCTTCATCAGCTTACGACACCCCTGACTTCGATGGGGATCACTATCTAATCACT 287  
DB 2090 eSerPheSerLeuGlnPheSer-----SerAsnGlyLeuHisTrp-----HisAsp 2106  
QY 286 CCAGAGAGCTCTGCGCATCTCCCTTGAAGGCTTCTCCACATGGCTCAGCTGCTCC- 231  
DB 2106 YTrpGAspLeuLeuPro-----GlylleuProLeuProGlyValSerPro 2122  
QY 230 -----AGTGTCT-----GTTGTGAC 212  
DB 2122 LAGlnGlyArgTrpGlyGlnGlnProThrMetProPheCysGlyPheHisSerLeuCysP 2142  
QY 211 CACAAAGCCCGACACT-----GGGATGCTCTTAC 180  
DB 2142 roGlnGlyProSerSerValProGlnGlyHisGlyLeuHis 2155  
RESULT 6  
ID 014687 PRELIMINARY; PRT; 4957 AA.  
AC 014687;  
DT 01-JAN-1998 (TrEMBLrel. 05, Created)  
DT 01-JAN-1998 (TrEMBLrel. 05, Last sequence update)  
DT 01-MAR-2002 (TrEMBLrel. 20, Last annotation update)  
DE ALR.  
GN ALR.  
OS Homo sapiens (Human).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
OC Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.  
RX NCBI\_TaxID=9606;  
RN (1)  
RP SEQUENCE FROM N.A.  
RA MEDLINE=97388474; PubMed=9247308;  
RA Prasad R., Zhadanov A.B., Sedkov Y., Bullrich F., Druck T.,  
RA Rallapalli R., Yano T., Alder H., Croce C.M., Huebner K., Mazo A.,  
RA Canani E.,  
RT "Structure and expression pattern of human ALR, a novel gene with  
RT strong homology to ALR-1 involved in acute leukemia and to Drosophila  
RT trihoax.";  
RL Oncogene 15:549-560(1997).  
CC -1- SIMILARITY: CONTAINS 1 SET DOMAIN.  
DR EMBL: AF010404; AAC51735.1; -  
DR InterPro: IPR003889; FYrich\_C.  
DR InterPro: IPR003888; FYrich\_N.  
DR InterPro: IPR000910; HMG\_12\_Box.  
DR InterPro: IPR003616; PostSET.  
DR InterPro: IPR002965; P-rich\_exten.  
DR InterPro: IPR001822; Recombinase.  
DR InterPro: IPR001214; SET.  
DR InterPro: IPR001965; Zn\_PHD.  
DR Pfam: PF00628; PHD; 3.  
DR Pfam: PF00856; SET; 1.  
DR PRINTS: PR01217; PRICHEXTENGN.  
DR SMART: SM00542; FYRC; 1.  
DR SMART: SM00541; FYRN; 1.  
DR SMART: SM00398; HMG; 1.  
DR SMART: SM00249; PHD; 4.  
DR SMART: SM00508; PostSET; 1.

DR SMART; SM00317; SET; 1  
 DR PROSITE; PS00398; RECOMBINASES\_2; UNKNOWN\_1  
 DR PROSITE; PS0280; SET; 2  
 SQ SEQUENCE 4957 AA; 531840 MW; 102652E1419CE8D CRC64;

## Alignment Scores:

Pred. No.: 0.00766 Length: 4957  
 Score: 127.50 Matches: 106  
 Percent Similarity: 33.97% Conservative: 36  
 Best Local Similarity: 25.36% Mismatches: 149  
 Query Match: 4.98% Indels: 128  
 DB: 4 Gaps: 22

US-09-989-919-15 (1-1397) X 014687 (1-4957)

QY 1219 CTTTCATGGAAGCAATGAG-----GCAGAAAGCCCCGGGCC 1181  
 DB 1765 LcnglyleuargProglngluProProProalaglAlaLeuAlaProserProProser 1784  
 QY 1180 CACCCC-----CGCCCCAGTCTGGCTCAACCAACAC-----CTGGCTC 1142  
 DB 1785 HlsProaspIleheargProglYserTyThrasProTyAlaglInProleuthr 1804  
 QY 1141 CCTGACCCAGTCTGGCTCATCTGAGTGCACCTTGCCCT----- 1096  
 DB 1805 ProargProglIn-ProProProProglInserCySAlaLeuProProargserleupr 1824  
 QY 1095 ----GACCCCTGGGCTGGCTGGAGTGTCTCTGGAGAGTCAAGATACCTCCCC 1040  
 DB 1824 oSerAapProheSerArgValProValSerProglInserSerSerInser 1844  
 QY 1039 AGGTACAGATCAACCCACATCTGAGACATCCCGCAAGTCTCCGTGCAGACCTAAG 980  
 DB 1844 OleuthrProargProleuser-----AlaglAlaPheCyProser----- 1858  
 QY 979 CTGAGAAACACTCCAGATGATGTCTCCACAGACCCCGCTGCTGTGCAGT 920  
 DB 1859 ----ProValThProargPhe-----GlnserProaspPro----- 1869  
 QY 919 TGCAGACAGACGACGAGGGGCTTACCTACATCTCACTATTGACACAGATGGG 860  
 DB 1870 ----TyrSe 1871  
 QY 859 CTGCCACGAGAACCTGCCAGAGAGGAGG-----CACAGCCCCAGG 818  
 DB 1871 rArgProProserArgProglInserArgaspProheAlaProleuHlsYProProAr 1891  
 QY 817 GGGAGACTAGCTGAGCTGTCTTACAGCCCAAGGGTTGAGAGGAGT-----CATAC 764  
 DB 1891 gProglInProProglInValAlaAlaheLys-----AlaglYSerleuAlaHlsH 1907  
 QY 763 CCCCAG--ATTCCTGAGCTAGAGGACGACAGCACTCCCATAGCTAG--GCTGTAAACTCAG 708  
 DB 1907 rSerleuGlyAlaglYglyPheProAlaAlaLeuProAlaglYProAlaglYLeuLeu-- 1926  
 QY 707 ATTTCACATTGGCATTATTCACCTTGTCTGTCTGATCATCTGTCTGTGGAGATGT 648  
 DB 1927 ----HsAlaLysValProserGlyInProProanPheValArgSerProglYh 1944  
 QY 647 AGCTTGTGTGACGGGTTCCAGGTGACATGACAGTGCAGTCAAGTCACT----- 593  
 DB 1944 rglYAlaPheValaglYThrProser-----PromerArgPhehPhehr 1959  
 QY 592 ----CCCTCACAGGAGCCAGCCAGCTG 570  
 DB 1959 oglAlaValaglYgluInProserleuLysProProValProglInProglYleuProProPr 1979  
 QY 569 CTGAGGATAGTCACTGTGT-----CAGCCAGCCCTCAGATAGGTCAGCCAGAGCCCTTAC 513  
 DB 1979 ohlsGlyIleAenSerHlsPheglYProglYProThrleuGlyLysProglInserThras 1999  
 QY 512 AATATGTGCTTGTGTGGATCAGGAGACACTGGCTCCATCCAGTCTTAAAAAGACCG 453

DB 1999 nTyThrValAlaThrclyAsn-----PheHlsPro-----Se 2010  
 QY 452 CTGCCGACCTGGGGGACAGAGGTGAGAGTAACAGGCTGATGTTCTTCTTGAGAG 393  
 DB 2010 rglYserProleuaglYProserSerGlySerThrclyglYInserTyrglyLeuser-Pro 2030  
 QY 392 CAAGACATTCCTCCCAATTGATGCTGCTTGAAGACAGCTATACCTCTTCATCACT 333  
 DB 2030 euArgProProserValLeu----- 2036  
 QY 332 CAGCACCCCTCTGACTTCCGATGGGAGATCACTATCATCAACTCCAGAGAGTCTG 273  
 DB 2037 ----ProProProAlaProasp---GlySerleuProTyrr---LeuserHlsGlyAla 2053  
 QY 272 CCATCTCCTTGGAGAGGCTTCTCCACT-----G 243  
 DB 2053 ergInArgserglYIleThrSerProValgluLysArggluLysProglYthrglymeG 2073  
 QY 242 GCTACCTGTCAGAGTGTCTGTGTGTGACACCAAGGCCGACACTGGGATGGCTCT 183  
 DB 2073 lYserSerleuAlaThrAlaGluleProglYthrglnaspProglYmetSerGlyLeus 2093  
 QY 182 CACAGACATCCACAGGCTCAGCCAGGTTGAACACTGAAGAGGAGGAGCTC 133  
 DB 2093 ergInThrclyleuGluLysglInArgglInArgglInArgleuArggluleu 2109

RESULT 7  
 ID 014686 PRELIMINARY; PRT; 5262 AA.  
 AC 014686;  
 DT 01-JAN-1998 (TREMblrel. 05, Created)  
 DT 01-JAN-1998 (TREMblrel. 05, Last sequence update)  
 DT 01-MAR-2002 (TREMblrel. 20, Last annotation update)  
 DE ALR.  
 OS Homo sapiens (Human).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.  
 OX NCBI\_TaxID=9606;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE=97388474; PubMed=9247308;  
 RA Prasad R., Zhadanov A.B., Sedkov Y., Bullrich F., Druck T.,  
 RA Rallapalli R., Yano T., Alder H., Croce C.M., Huebner K., Mazo A.,  
 RA Canaan E.  
 RT "Structure and expression pattern of human ALR, a novel gene with  
 RT strong homology to ALR-1 involved in acute leukemia and to Drosophila  
 RT trithorax."  
 RL Oncogene 15:549-560 (1997).  
 CC - SIMILARITY: CONTAINS 1 SET DOMAIN.  
 DR EMBL, AF010403; AAC51734.1, --  
 DR InterPro; IPR003889; FYRIC.1, --  
 DR InterPro; IPR003888; FYRIC.N.  
 DR InterPro; IPR000910; HMG\_12\_box.  
 DR InterPro; IPR003616; PostSET.  
 DR InterPro; IPR002965; P-rich\_extensn.  
 DR InterPro; IPR001822; Recombinase.  
 DR InterPro; IPR001214; SET.  
 DR InterPro; IPR001965; ZnF\_PHD.  
 DR Pfam; PF00628; PHD; 5.  
 DR Pfam; PF00856; SET; 1.  
 DR PRINTS; PR01217; PRICHEXTENSN.  
 DR SMART; SM00542; FYRC; 1.  
 DR SMART; SM00541; FYRN; 1.  
 DR SMART; SM00398; HMG; 1.  
 DR SMART; SM00249; PHD; 7.  
 DR SMART; SM00508; PostSET; 1.  
 DR SMART; SM00184; KING; 3.  
 DR SMART; SM00317; SET; 1.  
 DR PROSITE; PS00398; RECOMBINASES\_2; UNKNOWN\_1.  
 DR PROSITE; PS0280; SET; 2.  
 SQ SEQUENCE 5262 AA; 564171 MW; 26B7C74CAD417E44 CRC64;

## Alignment Scores:

Pred. No.: 0.00774 Length: 5262  
 Score: 127.50 Matches: 106  
 Percent Similarity: 33.97% Conservative: 36  
 Best Local Similarity: 25.36% Mismatches: 149  
 Query Match: 4.98% Indels: 128  
 DB: 4 Gaps: 22

US-09-989-919-15 (1-1397) x 014686 (1-5262)

QY 1219 CTTTCATTGAACCAATGAG-----GAGAAAGCCCGGCC 1181  
 DB 2070 LeuGlyLeuArgProGlnGluProProAlaGlnAlaLeuAlaProSerProProSer 2089  
 QY 1180 CACCCC-----CGCCCGAGCTCCTGCTCAACCAACAC-----CTGCTC 1142  
 DB 2090 HisProAspIlePheArgProGlySerThrAspProTyrAlaGlnProLeuThr 2109  
 QY 1141 CCTGACCCCGAGCTTGGCTCATGCTGAGGTGACCTCTGCCCT----- 1096  
 DB 2110 ProArgProGln-ProProProProGlySerCysAlaLeuProProArgSerLeuP 2129  
 QY 1095 ----GACCCCTGGGCTGGCTGGAGGTCTCCCTGTCGGAGGTCAAGATACCTCCC 1040  
 DB 2129 oSerAspProPheSerArgValProValSerProGlnSerGlnSerSerGlnSerP 2149  
 QY 1039 AGGTACAGATACCCACATCTCGAGACATCCCGGCAAGTCTCTGTCACACCTAG 980  
 DB 2149 oLeuThrProArgProLeuSer-----AlaGlnAlaPheCysProSer----- 2163  
 QY 979 CTGAGAAACACTCCAGATGATGTCCTCCAGCAGCCCGCTGCTGTCGATG 920  
 DB 2164 ----ProValThrProArgPhe-----GlnSerProAspPro----- 2174  
 QY 919 TGCAGACAGCAGCAGAGGGCTTACCTACATTCCTACACTATTGAACACAGTATGGG 860  
 DB 2175 ----TyrSer 2176  
 QY 859 CTGCCACGGAACCGTCCCGAGGAGGG-----CACAGCCCGCCAGT 818  
 DB 2176 rArgProProSerArgProGlnInerArgAspProPheAlaProLeuHisLysProProAr 2196  
 QY 817 GGGAGACTAGCTGAGCTGTACAGCCACAGAGGTGAGAGGAGT-----CATAC 764  
 DB 2196 gProGlnProProGluValAlaPheLys-----AlaGlySerLeuAlaHisThr 2212  
 QY 763 CCCCATG--ATTCTGAGCTAGAGAGCCAGAGTCCCATAGCTAG--GCTGTAACTCAC 708  
 DB 2212 rSerLeuGlyAlaGlyIlePheProAlaLeuProAlaGlyProAlaGlyIleLeu-- 2231  
 QY 707 ATTTTAACATTGGCATATTATGACGTTTGTCTGTCACATCTGTCTGGAGATGT 648  
 DB 2232 ----HisAlaLysValProSerGlyGlnProProAspPheValArgSerProGlyTh 2249  
 QY 647 AGCTTGTGTGACGGGCTCCAGGTGACATGACGTCTCCAGTGAACACTACT----- 593  
 DB 2249 rGlyAlaPheValGlyThrProSer-----ProMetArgPheThrPheP 2264  
 QY 592 -----CCCTCCACAGCCAGCCAGCTG 570  
 DB 2264 oGlnAlaValAlaGlyIleProSerLeuLysProProValProGlnProGlyLeuProProP 2284  
 QY 569 CTGAGATAGTACGCTGCT--CAGCCGAGCCCTCAGATAGTTCAGCCAGAGCCCTTAC 513  
 DB 2284 oHisGlyIleAsnSerHisPheGlyProGlyProThrLeuGlyLysProGlnSerThrAs 2304  
 QY 512 ATATGAGCTTGTGGGATCAGGACCACTTGAGCTCCAGCTGTGAAAAAGCCAG 453  
 DB 2304 nTyrThrValAlaThrGlyAsn-----PheHisPro-----Se 2315  
 QY 452 CTGCCACCTGGGGGCGAGAGGTGAGAGTAAAGGGCTGATGTTTCTTTCTGGAGG 393

DB 2315 rGlySerProLeuGlyProSerSerGlySerThrGlyIleSerTyrGlyLeuSer-ProL 2335  
 QY 392 CAGACCATTTCCCAATTGACTGCTTGAAGACAGTAACTCTCTCATCAGCT 333  
 DB 2335 euArgProProSerValLeu----- 2341  
 QY 332 CAGACCCCTCTGACTTCCGATGGGGATCACTATCTATCAACTCCAGAGAGTCTG 273  
 DB 2342 ----ProProProAlaProAsp--GlySerLeuProTyr--LeuSerHisGlyAlaS 2358  
 QY 272 CCATCTCCCTTGAAGAGCTTCTCCACT-----G 243  
 DB 2358 rGlnArgSerGlyIleThrSerProValGlnLysArgGlnLysProGlyThrGlyMetG 2378  
 QY 242 GCTCAGCTGTCAGAGGTGCTGCTGTCGATCAGCCAGAAAGCCGACACTGGGATGCT 183  
 DB 2378 LysSerLeuAlaThrAlaGluLeuProGlyThrGlnAspProGlyMetSerGlyLeuS 2398  
 QY 182 CACAGACATCCACAGCTCAGCCAGGTTGAACACTGMAAGAGAGCAGCTC 133  
 DB 2398 rGlnThrGluLeuGlnLysGlnArgGlnArgGlnArgLeuThrGluLeu 2414

RESULT 8  
 ID Q9N1P0 PRELIMINARY; PRT; 818 AA.  
 AC Q9N1P0;  
 DT 01-OCT-2000 (Tremblrel. 15, Created)  
 DT 01-OCT-2000 (Tremblrel. 15, Last sequence update)  
 DT 01-MAR-2001 (Tremblrel. 16, Last annotation update)  
 DE Submaxillary mucin (fragment).  
 GN BSM1.  
 OS Bos taurus (Bovine).  
 OC Eukaryota; Metazoa; Chordata; Vertebrata; Eumleostomi;  
 OC Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovidae;  
 OC Bovidae; Bovinae; Bos.  
 OX NCBI\_TaxID=9913;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=BRED ANGUS;  
 RX MEDLINE=20223253; PubMed=10759843;  
 RA Jiang W., Gupta D., Gallagher D., Davis S., Bhavanandan V.P.;  
 RT "The central domain of bovine submaxillary mucin consists of over 50  
 RT tandem repeats of 329 amino acids: chromosomal localization of the  
 RT BSM1 gene and relations to ovine and porcine counterparts.";  
 RL Eur. J. Biochem. 267:2208-2217(2000).  
 DR EMBL, AF178428; AAF67279.1; .  
 FT NON\_TER 1  
 FT TER 818  
 SQ SEQUENCE 818 AA; 73014 MW; B5BB4F84F66F86B CRC64;

Alignment Scores:  
 Pred. No.: 0.0106 Length: 818  
 Score: 124.50 Matches: 129  
 Percent Similarity: 34.41% Conservative: 41  
 Best Local Similarity: 26.11% Mismatches: 162  
 Query Match: 4.86% Indels: 162  
 DB: 6 Gaps: 22

US-09-989-919-15 (1-1397) x Q9N1P0 (1-818)

QY 1191 GGGCCCGGCGCCAGCCCGCCAGAGTCTGGCTCAACACACCTGCTCCGACCCCA 1132  
 DB 140 GlySerSerAlaThrAlaGlyProAlaValAlaGlySerGlnThrThrValGlnValSerGly 159  
 QY 1131 GTCCTTGCTCAGCTGAGGTGTCACCTTGCCTGACCCCTGAGCTGGCTGGAGTGT 1072  
 DB 160 AlaThrGlyThrAspValLeuLysSerGlyAlaSerLeuProValSerGlyAlaVal 179  
 QY 1071 TCCCTCTGTCGAGAGTCAAGATAGCTCCCAAGTCAAGATCAACCACTCTGAGAC 1012  
 DB 180 SerProGly-----SerSerProGlyArgSerArgAlaThr----- 192  
 QY 1011 ATCCCCCGCAAGTCTCTGTCACAGACTAAGCTGAGAGAACACTCCAGATGATGTC 952

```

Db 193 -----ValSerGlyGlnGlySerGln----- 199
Qy 951 CCCAGCAGCCGCCCGCTGCTGGCAGATGACAGCAGCAGAGGGCTTAGCT 892
Db 200 ProthValAlaLeuSerGlyAlaThnGlyThrSerAlaGlyProSerGlyThrArgSer 219
Qy 891 ACATTCTCACATTTTGAACACAGTATGGGGCTGCCAGCAGACCGTCCAGGAGGG 832
Db 220 Thr-SerSerAlaIleProAlaThrProGlySerThrThnGlyArgAlaAlaGlyAlaG 239
Qy 831 CACA----- 828
Db 239 YHrProAlaValAlaSpSerGlnThrAlaSerLeuProAlaAlaAlaArgProthral 259
Qy 827 -----GCCCCAGTGGGAGACTAGCTGAGCTTGCTTACAGCCCA 787
Db 259 AleuGlyProGlyThrSerAlaProSerGlyGlnThrSer--GluSerArgSerSerVal 278
Qy 786 CAGGTTGGAGAGGAGCAGTACATACCCCATGATTCCTGAGTACAGCAGCAGTCCCAT 727
Db 279 ProGly-----GlySerGlnThrThnGlnProGlyAlaGlySerGluSerProthr 296
Qy 726 GCTA-----GGCTGGTAAACTCAGCATTTTAA 700
Db 297 -LeuSerProGlyAlaThrArgThrThralaLeuArgGlySerGlnThrArgValProse 316
Qy 699 ATTGGCATTATTCAGCTTGTCTGCTGACATC----- 665
Db 316 rThnGly-----ValSerGlyLeu-ProGlySerThnGlnGlySerAlaAlaThnGlyG 335
Qy 664 -----TGCTGCTGGAGATGATTAATTCCTGAGGGGTTCCAGGTGACATGACAG 613
Db 335 lysSerGlyAlaGlySerGlyProthralaProValSerGlyGlnThrArgThrSerVal 355
Qy 612 TGCTGACAGTACATCCCTCCACAGCAGCCAGCTGCTGAGTACAGTCACTG 553
Db 355 leSerGlyThrAlaSerVal-----ProValSerGlyAlaProValThrProGlySerSerAlaG 374
Qy 552 GTCAGCGC---AGCCTCAGATAGTACAGCCAGAGGCTTCACATATGCTGCTTG--- 500
Db 374 lysSerGlyAlaProGlyThrGlyGlyProGlySerGlnThralaSerProLeuSerG 394
Qy 499 -----TTGGATCAGGAGAC----- 485
Db 394 lyaAlaAlaGlyThrSerAlaThrGlySerGlyThrSerIleProProSerGlyAlaProV 414
Qy 484 --ACTGCTCCATCAGTCTTG-----A 463
Db 414 alThrProGlnProProLeuIleSerThrGlyAlaSerAlaGlyProProAlaSerSerG 434
Qy 462 AAAAGACGAGCTGCCAGCTGGGGGAGAGGTAGAGTACAGGGCTGATTTCTC 403
Db 434 luserThrValThrLeuProGly-----AlaThnGlyThrAspValLeuA 449
Qy 402 TTCTGAGAGGAGAACCATTCCTCAATTGATGCTTGAACACGATATACCTCTC 343
Db 449 rgsSerGly-----ThrSerLeuProValS 457
Qy 342 CTCATCAGCTCAGACCCCTCTGACTTCGATGGGGATCATATCATCACTCCAG 283
Db 457 erglyGlyAlaValThrProAla-ProSerProGlyGlySer-----SerAlaThr 473
Qy 282 AGAGTCTCT-----GCCATCTCCCTTGGAGAGGCTTCTCCACATGGCTCACT 235
Db 474 AlaGlyProGlyValGlySerAlaThrThrValGlnAlaSerGlyAlaThnGlyAlaSp 493
Qy 234 GTCAGAGTGTCTGTTGTGTATCAACAAAGCCGAGACTGGGATGGCTCTCAGACACA 175
Db 494 Val-----LeuArgSerGlyThr 499
Qy 174 TCACAGCTCAGCAGGTTGAACATGAAAGAGGAGCTCCGCTGCT----- 126

```

```

Db 500 SerLeu-ProValSerGlyValAlaValaSerProGlySerSerProGlyArgSerGlyAl 519
Qy 125 -CGTAGATGCGCCAGCAGCGGATGCTTCC-----GGGAGATGGG 85
Db 519 aThrAlaValSerSerGlnGlySerGlnProthrValAlaLeuSerGlyAlaThnGlyTh 539
Qy 84 CTGCTGGATATCAGTCTGATCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 25
Db 539 rSerValGlyProSerGlyThrArgPheSer-----SerSerAlaIleProAl 555
Qy 24 CCGTCCCGT-----ACAGTGACACA 3
Db 555 aThrProGlySerThrThnGlyArgAla 564

RESULT 9
Q9EPX2 ID Q9EPX2 PRELIMINARY; PRT; 1280 AA.
AC Q9EPX2;
DT 01-MAR-2001 (TREMBlrel. 16, Created)
DT 01-MAR-2001 (TREMBlrel. 16, Last sequence update)
DT 01-JUN-2002 (TREMBlrel. 21, Last annotation update)
DE Papilin.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=CD-1;
RA MEDLINE=20530499; PubMed=11076767;
RA Kramarova I.A., Kusache-Gullberg M., Kramer J.M., Ackley B.D.,
RA Kramarova A.A., Kusache-Gullberg M., Kramer J.M., Ackley B.D.,
RA Stieren A.L., Prockop D.J., Fessler J.H.;
RT "Papilin in development: a pericellular protein with a homology to the
RT ADAMTS metalloproteinase."
RL Development 127:5475-5485(2000).
DR EMBL; AF314171; AAC641980.1; -.
DR HSSP; P12111; 2KMT.
DR InterPro; IPR000183; Decarboxylase2.
DR InterPro; IPR003598; Ig.
DR InterPro; IPR003598; Ig_C2.
DR InterPro; IPR003600; Ig_C2.
DR InterPro; IPR003600; Ig_1like.
DR InterPro; IPR003606; Ig_1like.
DR InterPro; IPR002223; Kunitz_BPTI.
DR InterPro; IPR000884; TSP1.
DR Pfam; PF00047; Ig_1; 3.
DR Pfam; PF00014; Kunitz_BPTI; 1.
DR Pfam; PF00090; tsp_1; 5.
DR PRINTS; PR00759; BASICPTASE.
DR ProDom; PD000222; Kunitz_BPTI; 1.
DR SMART; SM00409; Ig; 3.
DR SMART; SM00408; IGC2; 3.
DR SMART; SM00410; Ig_1like; 2.
DR SMART; SM00131; KUF; 1.
DR SMART; SM00209; TSP1; 5.
DR PROSITE; PS00280; BPTI_KUNITZ_1; 1.
DR PROSITE; PS0279; BPTI_KUNITZ_2; 1.
DR PROSITE; PS00879; ODR_DC_2_2; UNKNOWN_1.
DR PROSITE; PS50092; TSP1; 3.
DR KX Immunoglobulin domain; Serine protease inhibitor.
SQ SQUENCE 1280 AA; 138824 MW; AE287705B561AF30 CRC64;

Alignment Scores:
Pred. No.: 0,0115 Length: 1280
Score: 124,50 Matches: 128
Percent Similarity: 30,348 Conservative: 41
Best Local Similarity: 22,988 Mismatches: 183
Query Match: 4,93% Indels: 206
DB: 11 Gaps: 31

US-09-989-919-15 (1-1397) x Q9EPX2 (1-1280)
Qy 3 TGCTGACCTGTACCGAGCGGAGTATCTGACAGATCCACGAGCAGACGATACCG 62

```

```

Db 375 CysSerValSerCysGlyGlyValGlnSerArgSerValTyrCysIleSerSerAsp 394.
QY 63 AGTACCATGATATCCAGACAGACACATCCCCAGAGAACTACCGCTGGCCATCT 122
Db 395 GlyThrGlyValGlnGlnValAlaGlnGlnThrGlnCysAlaGlyLeuAlaGlyLysPro 414
QY 123 ACCACCAAGGAGCTGCTCTCTTACG--TGTTCAACCTG--CTGAGCTGTGGAT- 175
Db 415 ProThrThrGlnAlaCysAsnLeuGlnIleCysLeuValTrpSerValGluProTrpGly 434
QY 176 -----GTCGTGAGAGCCATGCCAGTGTGG----- 202
Db 435 GluCysSerValThrCysGlyThrGlyIleArgLysArgSerValThrCysArgGlyAsp 454
QY 203 -----GCCTTGTGGTCCACACCAAGCACCCTGGACAGCT 238
Db 455 GluGlySerProValHisAlaAlaAlaCysLeuLeuLysAspGlnProThrLeuThrGlu 474
QY 239 GAGCCAGTGGAGAGAGCCCTTCCAGAGGAGATGGACAGACCTCTCTGAGGTTGATAGAT 298
Db 475 ProCysValGlnGlnAlaCysPro--ValPheArgGlyGlnAlaTrpHis--ValGly 492
QY 299 AGTGATCCCCCATCGAAGTCAAGAGGGGCTCTGAGTGTAGAGAGAGATATACGTGT 358
Db 493 SerTrpSerLeuCysSerLysSerCysGlySerGlyIleArgArgGlnValAlaCys 512
QY 359 CTTCAAGGAGTCAAAATTAGGAGAAATGTCCTTCCCTCCAGAAAGAA----- 406
Db 513 ThrIleGlyProProGlyArgCysValAspLeuGlnSerSerLysProAlaGluMetGlu 532
QY 407 -----AACATCCAGCCCTGTATACCTC-----TCACCTCTGCCCCAG 445
Db 533 AlaCysAsnArgGlnProCysHisIleuProGlnGluValProSerIleGlnAspProArg 552
QY 446 TCGGCAGCT-----GGT 457
Db 553 ThrArgSerSerAspProArgIleLeuSerGlyProArgValSerProValSerAspGly 572
QY 458 CTTTTCAGAGCTGGATG-----GAGCCAGAGTGGT 487
Db 573 ArgGlnGlnGlnThrAlaProLeuGlnArgProArgAlaGlnSerAspProArgGlyGly 592
QY 488 CCCTGATCCCAAGACAGACATATATGAGGCC----- 521
Db 593 GlnAspProAsnLeuSerSerAlaGlyArgAlaProThrLeuGlnArgProProHisGly 612
QY 522 -----TCTGGCTGACCTATCTGAGGGCTGGCTGACAGCTGACTA 562
Db 612 nProProLeuArgProSerSerGly-----ProArgAspCys 624
QY 563 T-----CCTCAGCA 571
Db 624 sarGHisSerProHisGlyCysCysProAspGlyHisThrProSerLeuGlyProGlnTr 644
QY 572 GCTGGCTTGGCTGTGGAGGAGATGACTTGCACATGGCAGACACT-----GCATGTCACT 625
Db 644 pGlnGlyCysProLeuAlaGlyAlaSerCysLeuGlnSerArgTyrGlyCysCysProAs 664
QY 626 GGGAGCCCTGACAGAAAGCT----- 647
Db 664 pGlyValSerAlaAlaGlnGlyProGlnGlnAlaGlyCysThrArgSerHisGlySerHis 664
QY 648 -ACATCCCAAGACAGACATGTGACACAGACAAAGCTGATATATATGAGGTTAAAA 706
Db 684 pAsnThrGlyAsnArg-----ProGlySerArgAlaValAlaSerLysAsnProLys 701
QY 707 TGTGAGTTTACAGACCTAGTATGAGACTGTGGCTCTAGTCCAGGAATCATGAGGGGTA 766
Db 701 sIleHisGlnProGln-----AlaHisGlnGlyGluProse 713
QY 767 TGACTGCTCTCCAAACCTGTGGGCTGT-----AGCAAGCTCAGGCTAGTCTCC 817

```

```

Db 713 rGluCysArgSerSerArgPheGlyCysCysTyrAspAsnValAlaSerAlaAlaGlyPr 733
QY 818 CACTGGGGGCTGTGGCCCTCCCTGGAGCGGTTCCGTGGGAGGCCCATCATCTGTTCAA 877
Db 733 oleuGly-----GluGlyCysValGlyGlnProSerTyrAlaTyr-- 746
QY 878 TAGGTGAGATGTACTAAAGCCCTGCTGCTGCTGCTGCATCCACAGACAGCGGT 937
Db 747 -----ProValArgCysLeuLeuProSerAlaGlnIleSerCys 759
QY 938 GGGG-----GTCGCTGG-----GACATATCATTCGTGAGTGTCTCTACGCTT 962
Db 759 sGlyAspTrpAlaAlaArgTrpTrpPheValAlaSerValAlaArgCys----- 775
QY 983 AGGTGTGACAGAGACCTTGGGGGAGATGCTCCAGCA----- 1020
Db 776 -----AsnArgPheTrpTyrGlyGlyCysHis-GlyAsnAlaAsnAspAlaSerG 793
QY 1021 -----TGCGGTATCTGTACTGTGGGAGGCTATCTTGACCTCCGACAGGGGA- 1071
Db 793 lngGlnGlyCysMetAlaThrCys-----ArgGlyG 803
QY 1072 -CACTCCAGGCCACCCCA-----GGGTCAGGGGACAGAGTGCACACTTCAGCATGA 1123
Db 803 lngHisGlyProArgArgProGluAlaGlyAlaAlaGlyHisArgAlaHisVal----- 820
QY 1124 GCCAAGACTGGGTGACAGGAGCAGAGCTGTGGTTGACCAGACCTGGCGGGGCTGG- 1182
Db 821 -----AspGlyGlyGlnArg-----GlyProGlyGlyG 830
QY 1183 -----GCCGGGAGCTTCTGCTCATTTGCTTTCATATGAATAAG 1219
Db 830 lngGlnIleProAspTrpHisArgAlaGlyAlaThrIleProArgLeuProSerProSerG 850
QY 1220 CCTCA-----AGCAGCCAAACACAGGCTTTCCTCCCT 1251
Db 850 lYserProTrpArgArgGlnGlnGlnProAlaProGlyGluProPro 865

RESULT 10
ID 075441 PRELIMINARY; PRT; 669 AA.
AC 075441;
DT 01-NOV-1998 (TREMBLrel. 08, Created)
DT 01-NOV-1998 (TREMBLrel. 08, Last sequence update)
DT 01-MAR-2002 (TREMBLrel. 20, Last annotation update)
DE Latent transforming growth factor-beta binding protein 4
   (Fragment).
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominiidae; Homo.
OX NCBI_TaxID=9606;
RN (1)
RP SEQUENCE FROM N.A.
RX MEDLINE=98325059; PubMed=9660815;
RA Saharinen J., Taipale J., Morni O., Keski-Oja J.;
RT "Identification and characterization of a new latent transforming
RT growth factor-binding protein, LTRP-4."
RT J. Biol. Chem. 273:18459-18469(1998).
DR EMBL; AF054502; AAC39882.1; -.
DR HSSP; P35555; 1EMN.
DR InterPro; IPR000152; Asx hydroxyl.
DR InterPro; IPR001881; EGF-like.
DR InterPro; IPR001881; EGF CA.
DR InterPro; IPR002212; Fibrin-asso.
DR Pfam; PF00008; EGF; 6.
DR SMART; SM00179; EGF CA; 6.
DR SMART; SM00001; EGF-like; 2.
DR PROSITE; PS00010; ASX HYDROXYL; 6.
DR PROSITE; PS01186; EGF_2; UNKNOWN_2.
DR PROSITE; PS01187; EGF CA; 8.
KW Calcium-binding; EGF-like domain; Glycoprotein; Hydroxylation; Repeat.
FT
NON_TER 1

```

FT NON TER 669 669  
 SQ SEQUENCE 669 AA; 71702 MW; 2C76C4AB11428A4 CRC64;  
 Alignment Scores:  
 Pred. No.: 0.0127 Length: 669  
 Score: 123.50 Matches: 109  
 Percent Similarity: 30.97% Conservative: 31  
 Best Local Similarity: 24.12% Mismatches: 129  
 Query Match: 4.89% Indels: 184  
 DB: 4 Gaps: 27  
 US-09-989-919-15 (1-1397) x 075441 (1-669)  
 QY 110 TGCTGGCATTCACCAACGAGG-----AGTCCTCTTTCAGTTTCAC 157  
 DB 128 CysAlaProGlyThrProGlyProAlaGlyAlaSerCysLeu----- 142  
 QY 158 CTGGCTGAGGCTGTGATGTCGTGAGAGCCATGCCAGTGGGCTTTGTGTCACC 217  
 DB 143 -----AspValAspGluCysSerGluGluAspLeuSerGlnSerGlyIleCysThr 159  
 QY 218 AACACAGACCACTGAGCAGGTGAGCCAGTGGAGAACCCCTCCAGAGGATGGCAGA 277  
 DB 160 Asn---ThrAspGlySerPheGluCysIle-----CysProProGlyHisArgGly 176  
 QY 278 CCA---CTGTGAGGTTGATAGATGATGATCCCATCGGAAGTCAGAGGGGCTGCTGAG 334  
 DB 177 ProAspLeuAlaSerCysLeuAspValAspGlu----- 187  
 QY 335 GTGATGAGAGAGAGGTATACGTGCTTCAAGGAGCATCAATTAAGGATGGCTTGC 394  
 DB 188 ---CysArgGluArgGlyProAlaLeuGlySerGlnArgCysGlnAsnSerProGly 206  
 QY 395 TCACAGAA-----AGAACATCCAGCC----- 418  
 DB 207 SerTyrArgCysValAlaArgAspCysAspProGlyTyrHisAlaGlyProGluGlyThrCys 226  
 QY 418 ----- 418  
 DB 227 AspAspValAsnGluCysGluThrLeuGlnGlyValGlySerGlyAlaAlaLeuCysGluAsn 246  
 QY 419 -----TGTTACCTCTCACT----- 433  
 DB 247 ValGluGlySerPheLeuCysValCysProAsnSerProGluGluThrAspProMetThr 266  
 QY 434 -----CTGCCCCCAGGCTGGCAGCTGGTCTTTTCAAGACTGATGAGCCAACT 484  
 DB 267 GlyArgCysValProProArgThrSerAlaGlyMetPheProGly---SerGlnProGln 285  
 QY 485 GGTCCCTGATCCCAACAAGACACATATGTGAAGGCTCTGGCTGACCTATCTGAGGGCT 544  
 DB 286 AlaProAlaSer-----ProValLeuProAla 294  
 QY 545 CGG---CTGACCACTGACTATCTCAGACAGCTGGCTGGCTGTGAGGAGAGTACTTG 601  
 DB 295 ArgProPro-PropProLeuProAlaGArgProSerThrProArgGlnGlyProValGln 314  
 QY 602 CACTGGCAGACATGTCATCACTGGAGACCCCTGACAGACAAAGCTAACTCCAGACAG 661  
 DB 314 ySerGlyArgArgGluCysTyrPheAspThrAlaAla-----ProAsp-- 328  
 QY 662 ACAGATGTGACGAGCAAAAGCTGCAATATGCAAAATGTTAAATGTGAGTTTACACAG 721  
 DB 329 -----AlaCysAspAsnIleLeuAlaArgAsnValThr----- 339  
 QY 722 CTAGCTATGG---CACTGCTGGCTCCCTAGTCCAGGAATCAATGAGGGGTATGACTGCTCTC 778  
 DB 340 -----TlrglnGluCysCysThrValGlyGlnGlyTyrPoli----- 352  
 QY 779 CAACCTGTGGGCTGTAGAGCAAGCTAGCTAGTCCCACTGGGGGCTGTGCCCTCC 838  
 DB 353 -----SerGlyCysArgIleGlnGlnCys-----Pr 361

QY 839 CTGGACGGTTCGTRGGGAGCCCATCACTGTGTTCAATAGTGAATGACTTAA 898  
 DB 361 o-GlyThrGluThrAlaGluTyrGlnSerLeuCysPro----- 373  
 QY 899 GCCCTGTGCTGCTGCTGCTGACATGCCACAGAGCGGTG----- 939  
 DB 374 -----HisGlyArgGlyTyrLeuAlaProSerGlyAspL 385  
 QY 940 ---CGGCTGGTGGGACATTCATGCTGAGTGTCTCTCACTTAAGTCTGG---ACAG 994  
 DB 385 eusSerLeuArgArgAspValAspGluCysGlnLeuPheArgAspGlnValCysLysSerG 405  
 QY 995 GAGACTTGGCGGGGAGTGTCCAGATGTGGTGTATCTGTACTGGGAGGCTATCTCT 1054  
 DB 405 lValCysValAsnThrAlaProGlyTyrSerCysTyrCysSerAsnGlyTyrTyr- 424  
 QY 1055 GACCTCCGACAGGAGGACCTCCAGGC----- 1082  
 DB 425 -----HisThrGlnArgLeuGluCysIleAspAsnAspGluCysAla 439  
 QY 1083 ---CAGCCAGAGGCTCAGGGGACAGGTGCACA-----CCTCAGCAT----- 1121  
 DB 439 spGluGluProAlaCysGluGlyGlyArgCysValAsnThrValGlySerTyrHisCysT 459  
 QY 1122 ---GAGCCAGACTG-----GGGTCAAGGAGCAGGTGT----- 1151  
 DB 459 hrcCysGluProProLeuValLeuAspGlySerGlnArgArgCysValSerAsnGluSerG 479  
 QY 1152 ---GGTTGAGCCAGACTGTGGGCGGGGTGG 1181  
 DB 479 InsSerLeuAspAspAsnLeuGlnGlyValCysTlrg 489  
 RESULT 11  
 Q9ARH1 PRELIMINARY; PRT; 647 AA.  
 ID Q9ARH1  
 AC Q9ARH1  
 DT 01-JUN-2001 (TrEMBLrel. 17, Created)  
 DT 01-JUN-2001 (TrEMBLrel. 17, Last sequence update)  
 DT 01-MAR-2002 (TrEMBLrel. 20, Last annotation update)  
 DE Receptor protein kinase PERK1.  
 OS Brassica napus (Rape).  
 OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
 OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;  
 OC eurosids II; Brassicales; Brassicaceae; Brassica.  
 OX NCBI\_TaxID=3708;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RA Silva N.F., Goring D.R.;  
 RT "Characterization of a novel Brassica napus receptor protein kinase  
 RT PERK1."  
 RL Submitted (MAR-2001) to the EMBL/GenBank/DBJ databases.  
 CC -1- SIMILARITY: BELONGS TO THE SER/THR FAMILY OF PROTEIN KINASES.  
 DR EMBL; AY028699; AAK21965.1; -  
 DR InterPro; IPR000719; Euk\_pkinase.  
 DR InterPro; IPR002965; P-rich\_kinase.  
 DR InterPro; IPR002290; Ser\_thr\_kinase.  
 DR InterPro; IPR001245; Tyr\_pkinase.  
 DR Pfam; PF00069; pkinase; 1.  
 DR PRINTS; PR01217; PRICHEXTENSN.  
 DR PRINTS; PR00109; TYRKINASE.  
 DR Prodom; PD000001; Euk\_pkinase; 1.  
 DR SMART; SM00220; S\_TKc; 1.  
 DR SMART; SM00219; TYKc; 1.  
 DR PROSITE; PS00107; PROTEIN\_KINASE\_ATP; 1.  
 DR PROSITE; PS00108; PROTEIN\_KINASE\_DOM; 1.  
 DR PROSITE; PS00108; PROTEIN\_KINASE\_ST; 1.  
 KW ATP-binding; Kinase; Serine/threonine-protein kinase; Transferase.  
 SQ SEQUENCE 647 AA; 68482 MW; A69AEAC00D5CA203 CRC64;  
 Alignment Scores:  
 Pred. No.: 0.0156 Length: 647  
 Score: 122.50 Matches: 91  
 Percent Similarity: 31.28% Conservative: 36





US-09-989-919-15 (1-1397) x Q9UPX0 (1-1400)

QY 1191 GGGCCCGGCCCCACCC-----GCCCGAGTCTGGCTCAAAACACACCTGCTCCTGACC 1135  
 DB 954 GlyserylProgluProleuAlaProSerPro-----ValSer 966

QY 1134 CCACTTGGCTATGCTGAGGTGTGACCTTGGCCCTGACCCCTGGGCTGGCTGGGA 1075  
 DB 967 ProThrPhePro-----ProSerSerProSerAspTrpPro-----GlnGlu 980

QY 1074 GTGTCCCTGCTCGGAGGATGAGATAGCTTCCCAAGTACAGATACACCCACATCC--- 1018  
 DB 981 ArgSerProGlyGlyHisSerAspGlyAlaSerProArgSerProValProThrIleu 1000

QY 1017 -----TGAGACATCCCGCCCAAGTCTCTGTCCAGACTAAGC 979  
 DB 1001 ProGlyLeuArgHisAlaProTrpGlnGlyProArgGly-----ProProAspSer 1017

QY 978 TGAGAGAACACTCCACATGATTTGTCCCAAGCAG-----CCC 940  
 DB 1018 ProAspGlySerProLeuThrProValProSerGlnMetProTrpLeuValAlaSerPro 1037

QY 939 CCAACGGCTGCTGGCATGTGACGACGACGACGAGGGCTTATAGCTTACATTTCTCAC 880  
 DB 1038 GluProPro-----GlnSerSerProThrProAlaPheProLeuAlaAlaSer 1053

QY 879 TATTGAACACAGTATGAGGCTGCCA-----CGAACCTCCCAAGGAG 835  
 DB 1054 TyrAspHisnGlyLeuSerGlnProProLeuProGlyAlaGHisLeuProGlyPro 1073

QY 834 GGGCACAGCCCC---CAGTGGGAGACTAGCTTGTGCTTACAGCCCAAGGGTTGG 778  
 DB 1074 GlyGlnGlnProGlyProTyrGly-----ProGlnGlnAlaSerSerProAla--- 1089

QY 777 AGAGGCACTATACCCCATGATTCCTGACTAGACCCAGACAGTCCCATAGTAGCTG 718  
 DB 1090 ArgGly----- 1091

QY 717 GTAACTCACATTTTAACATTGGCATTTATGACGTTTCTCTGTCACATCTGTCTGT 658  
 DB 1092 -----IleSerHisIleValIleThrPheAlaProLeuLeuSer--- 1103

QY 657 CTGGAGTGTAGCTTGTCTGACAGGGTCCAGAGTACATGACAGTGTCTCCAGTGAAG 598  
 DB 1104 -----Asp 1104

QY 597 TCACTCCCTCCACAGGACGCCAGCTGTGAGATAGTACGTGTACGCCGACCTC 538  
 DB 1105 AsnValProGlnIleProGluProProThiGlnIleSerGlnSerAsnValIlePheVal 1124

QY 537 AGATAGGTACAGCCAGAGCCCTTACATATGTGTTGTGGATCAGAGCACTTGGC 478  
 DB 1125 GlnAspHisSerIle-----PheTrp-TyrIleProHisLeuSer 1137

QY 477 TCCATCC-----AGCTTGAAGAAAGACCACTGCCACCTGGGGGGGAGAGGTGAG 427  
 DB 1137 TArgAspGlnAlaIleAlaLeuLeuIleAspIleAspPro----- 1150

QY 426 AGTAAACAGGGCTGAGTGTCTCTTCTGAGGACAGACATCTCCCTAATTGACTG 367  
 DB 1151 -----GlyAlaPheLeuIle---ArgAspSerHisSer--- 1160

QY 366 CCTTGAAGACAGTATACCTCTCTCTCATACCTCAGACACCCCTGTGACTTCGATGG 307  
 DB 1161 PheGlnGlyAlaIleTrp-GlyLeuAlaLeuIleValAlaIleThrPro----- 1174

QY 306 GGATCACTATATCAACCTCCAGAGAGGTCTGTGACATCTCTTGGAGAGGC----- 254  
 DB 1175 -----ProProSerAlaGln-----ProTrpIleGlyAspProVal 1186

QY 253 -----TTCCTCCACATGGCTC-----ACCTGTCCAGAGGT-----CTGGTTG 217

DB 1186 aIGluGlnLeuValArgHisPheLeuIleGluThrGlyProIleGlyValIleIleYSG 1206

QY 216 GTGACCACAAAGCCCGACACTGGCC-----ATGGCTCTCACAGACATCCACAGCTCA 163  
 DB 1206 LysAspProSerGlnProTyrPheGlySerLeuSerAlaLeuValSerGlnHisSerIle 1226

QY 162 GCCAGGT-----GAACACTGAA 145  
 DB 1226 eIFProIleSerLeuProCysCysLeuArgIleLeuSerIleAspProLeuGluGluThrP 1246

QY 144 AGAGGACACTCCCGTGTGTGAGATGCGCAGACGGCTAGCTTCTCTGGGAGATGG 85  
 DB 1246 roGlu---AlaProValProThrAsnMetSerThrAlaAlaAspLeuArgGlnGlyA 1265

QY 84 CTGTCTGGGATACACTGTACTC 62  
 DB 1265 IaAlaCysSerValLeuIleu.1272

RESULT 13  
 Q9UPX0  
 ID Q9UPX0 PRELIMINARY; PRT; 763 AA.  
 AC Q9UPX0;  
 DT 01-MAY-2000 (TREMBLrel. 13, Created)  
 DT 01-MAY-2000 (TREMBLrel. 13, Last sequence update)  
 DT 01-DEC-2001 (TREMBLrel. 19, Last annotation update)  
 DE KIAA1030: protein (Fragment).  
 GN KIAA1030.  
 OS Homo sapiens (Human).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
 OX NCBI\_TaxID=9606;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC TISSUE=BRAIN;  
 RX MEDLINE=9397452; PubMed=10470851;  
 RA Kikuno R., Nagase T., Ishikawa K., Hirosewa M., Miyajima N.,  
 RA Tanaka A., Korani H., Nomura N., Ohara O.;  
 RT "Prediction of the coding sequences of unidentified human genes. XIV.  
 RT The complete sequences of 100 new cDNA clones from brain which code  
 RT for large proteins in vitro.";  
 RL DNA Res. 6:197-205 (1999).  
 DR EMBL; AB028953; BAA82982.1;  
 DR InterPro; IPR003961; FN\_III.  
 DR Pfam; PF00041; fn3; 1.  
 DR SMART; SM00060; FN3; 1.  
 FT NON TER 1  
 SQ SEQUENCE 763 AA; 82322 MW; B9756B33F2F2B619 CRC64;

Alignment Scores:  
 pred. No.: 0.0245 Length: 763  
 Score: 120.50 Matches: 84  
 Percent Similarity: 31.40% Conservative: 35  
 Best Local Similarity: 22.16% Mismatches: 134  
 Query Match: 4.71% Indels: 126  
 DB: 4 Gaps: 16

US-09-989-919-15 (1-1397) x Q9UPX0 (1-763)

QY 1182 CCCACCCCGCCCGCCAGTCT-----GGCTCAAC 1153  
 DB 414 ProLeuProThrGluGluGlyProPheGlyHisPProThrIleProGluGluIleAsn 433

QY 1152 CACACTGCTCCTGACCCCAAGCTTGGCTCATGCTGAGGTGACACCTCTGCCCTGAC 1093  
 DB 434 AlaSerAsnSerThrIleuPro---LeuThrGlnThrProThrIleGlyIleArgSerProGlu 452

QY 1092 CCCTGGGCTGGCTCGGAGAGTCCCTGTCGGAGGTGACAGATACCTCCACAGGTACA 1033  
 DB 453 ProTrp-----GlyArgProGluPheProPheGlyGlyLeu 464

QY 1032 GAATCACCAACATCTGAGACATCCCGCCCAAGTCTCTGTCCAGACCTAAGCTGAG 973  
 DB 465 GluThrPro---AlaIleMetPheProHisGlnLeuProProCysAspValProGluSer 483



QY 972 AACACTCAGAGATGTTGCCAGCAGCCCGCTGTCATGTGAGCA 913  
DB 484 LeuGlnProLysAlaGlyLeuProArgGlyLeuProProHisSerLeuGlnValProAla 503  
QY 912 GCAGCAGCAGGGCTTTAGCTACATTTCTACACTATTGAACAGATGATGGCTGCCCA 853  
DB 504 AlaTyrProGlyIleLeuSerLeuGlnAlaProLys-----GlyTyrAla 518  
QY 852 CGAAGCCGTCAGAGGGGAGCAGAGCCG-----CAGTGGAGAGCTAG 808  
DB 519 GlyLysSerProLysArgGlyProAlaProAlaAlaLysTyrGlnAspArg 538  
QY 807 CCGAGCTGCTTACAGCCAGCAGGTTGG-----AGAGCAGCTCATACC 763  
DB 539 PrometGlnProLeuValSerGlnGlyGlnLeuAlaHisThrSerGlnGlyMetGlyIle 558  
QY 762 CCCATGATTTCTGACTAGAGCAGCAGTCCCATAGCTAGGCTGTAACATCATTTT 703  
DB 559 ProValLeuProTyrProGlnProAlaGlnProGlyAlaHisGlyGlyProSerThrPhe 578  
QY 702 AACATT-----TGGCATTAATGACGTTTGTCTGTACATCTGTCTGTGGGA 652  
DB 579 GlyLeuAspThrArgTyr-----AGAGCAGCTCATACC 585  
QY 651 TGTAGCTTTGTCTGAGGGGTTCCAGTACATGAGTGTGCTGACAGTCACTC 592  
DB 586 -----GluPro 587  
QY 591 CCTCCAGAGCAGCAGCCAGCTGTCAGATAGTACGTGTGACCCAGCCTCAGATAG 532  
DB 588 GlnProArgProArgProSerProArgGlnAlaArgAlaGlnProSerLeuHisGln 607  
QY 531 GTGAGCAGCAGGGCTTACATATGTGTCTTTGGATAGAGACCATTTGCTGCATC 472  
DB 608 ValValLeuGlnProSerAlaGlyLeu-----SerProLeuThrGlnSerProLeuSerSer--- 625  
QY 471 CAGCTTTGAAAAAGACAGCAGCTGCGCAGCTGGGGGAGAGTGAAGAGAGGCTGG 412  
DB 626 -----ArgThrGlySerProGlnLeuAlaAlaArg----- 635  
QY 411 ATGTTTCTTTCTGAGAGCAGACCATTTCTCCCTAATTGACTGCTTGAAGACAGTA 352  
DB 636 -----AlaArgPro----- 638  
QY 351 TACCTCTCTCATCACTAGACACCCCTCTGACTTCCGATGGGGATCTATCTATC 292  
DB 639 -----ArgProGlyLeuLeuGlnGlnAlaGlnMetSerGluIle 651  
QY 291 AACCTCAGAGAGGTCTGCTCATCTCCCTTGAAGAGGCTTCT-----CCCAGTGGCTCA 238  
DB 652 ThrLeuGlnProProAlaAlaValSerPheSerThrGlySerThrProSerThrGlySer 671  
QY 237 CCGTCCAGAGTGTCTGTTGTGTGACCAAGAGCCGACAGCTGGGAGTGGCTCTCAG 178  
DB 672 ProSerGln-----SerSerArg-SerGlySerProSerTyrArg 684  
QY 177 A-----CATCCAGACCTTACAGCAGT 156  
DB 684 gProAlaMetGlyPheThrThrLeuAlaThrGlyTyrProSerProProProGly 702

OX NCBI TaxID=9606;  
RN (1)  
RP SEQUENCE FROM N.A.  
RA Daggett L.P., Lu C., Johnson E.C., Vellicelebi G.,  
RT "Cloning, Distribution and Functional Expression of the Human NMDAR2C  
RT Receptor";  
RL Submitted (NOV-1996) to the EMBL/GenBank/DBJ databases.  
DR EMBL; U77782; AAB71624.1; -  
DR HSSP; P19491; 1GR2.  
DR InterPro; IPR001320; Ion\_glu\_receptor.  
DR InterPro; IPR001311; SBP/glu\_receptor.  
DR Pfam; PF00060; 1lg\_chan; 1.  
DR SMART; SM00079; PBPe; 1.  
KW Receptor; signal.  
FT SIGNAL 1 21  
FT CHAIN 22 1236 N-METHYL-D-ASPARTATE RECEPTOR 2C SUBUNIT.  
SQ SEQUENCE 1236 AA; 134462 MW; E57JD3B826290A0FI CRC64;  
  
Alignment Scores:  
Pred. No.: 0.0266 Length: 1236  
Score: 120.50 Matches: 88  
Percent Similarity: 32.37% Conservative: 35  
Best Local Similarity: 23.16% Mismatches: 135  
Query Match: 4.71% Indels: 122  
DB: Gaps: 19  
  
US-09-989-919-15 (1-1397) x 015398 (1-1236)  
  
QY 1199 GGCAGAAAGCCCGGCCCGCCAGCCCG-----CCGAGTCTTG 1161  
DB 926 GlyArgArgAlaProProProSerProCysProThrProArgSerGlyProSerProCys 945  
QY 1160 GCTCAAGC-----ACAGCTGCTCCCTGAGCCCGAGCTTGG----- 1125  
DB 946 LeuProThrProAspProProProGlnProSerProThrGlyTyrPheProProAspGly 965  
QY 1124 -----CTCATGTGAGTGTGACCT-----CTGCCCTGAGCCCTGGG 1086  
DB 966 GlyArgAlaAlaLeuValArgAlaProGlnProProGlyArgProProThrProGly 985  
QY 1085 CTGGCTGAGAGTGTCCCTGTCGGAGGTCAAGATTAAGCTCCAGAGTACAGAAATCAC 1026  
DB 986 -----ProProLeuSerAspValSerArg-----ValSerArgArg 997  
QY 1025 CCACATCTGAGAGCATCCCGCCAGCATCTCCCTGTCAGACC-----TAAGCTGAGAGACA 969  
DB 998 Pro-----AlaTyrGlnAlaArgTyrProValArgThrGlyHisCysGlyArgHis 1014  
QY 968 CTCAGATGATGATTGCCCCAGCAGCCCGCCAGCTGTCGATGTCAGCAGCAG 909  
DB 1015 LeuSerAlaSerGlnArgProLeuSerProAlaArg----- 1026  
QY 908 CAGCAGGGCTTTAGTACATTTCTACACTATTGAACAGATGAGGGCTGCCACGGA 849  
DB 1027 -----CysHisTyrSerSerPheProArgAlaAspArgSerGlyArgPro 1041  
QY 848 ACCGTCCAGAGAGGGGACAGCCCGCCAGTGGGGAATAGCTGAGCTTGTCTTACAGCC 789  
DB 1042 PheLeuProLeuPheProGlnProProGlnLeuGln-----AspLeuProLeu-LeuGlyPhe 1060  
QY 788 CACAGGGTTG-----GAGAGCAGTCATACCCCATAGTCTGTCAGTACAGAGCAGCAGTC 732  
DB 1060 GlnGlnLeuAlaArgArgGlnAlaLeuLeuHisAlaAlaTyrPheAlaArgGlySerArg 1079  
QY 731 CCATAGCTAGGCTGTAAATCATCATTTTAACATTTTGAGATTATTCAGAGTTTCTCTG 672  
DB 1079 ----- 1079  
QY 671 TCACATCTGTCTGTGAGATGTAGCTTTGTCTGAGAGGGTTCCAGGTGATCATGCACT 612  
DB 1080 -----ProArg-HisAlaLeu 1084  
QY 611 GTGCAGTGAAGTCACTCCCTCCACAGGAGCCAGCCAGCTGTGAGATGATGACCTGG 552

```

Db      1084 rleupProserSeValAlaGluAlaPheAlaArgProser-----SerleuProAlaG 1102
QY      551 TCAGCCGAGCCCTCAGATAGTGCAGGAGGCGCTTCATATATGTGCTTGTGGGATC 492
Db      1102 ycsYthrglyProAlaCysAlaArgProAserglyHissEralaCysArgArgleuAlaG 1122
QY      491 AGGAGACACTTGGCTCCATCTTCGTAAGAAAAGACAGCTGCCGACCTGGGGGAGAG 432
Db      1122 naLaGlnSerMetCysleuProIleTyrArgGlu---AlaCysGlnGlnGlyGlnGln- 1140
QY      431 GTGAGAGGTAAACAGGGGTGATGTTCTCTTCTTGAGGCAAGACATTCCTCCATATT 372
Db      1141 -----AlaGlyAla 1143
QY      371 GACTGCTTGAAGACAGATATACCTCTCTCTATCATACCTGACAGCCCTCTGATCTCCG 312
Db      1143 apProAlatrgPlnHlsArgGlnHlsValCysLeuHlsAlaHlsAlaHlsleuProPhocy 1163
QY      311 ATGGGGATCATATATCATACCTCCAGAGAGTCCCTGCATCTCCCTTGGAAAGGCTT 252
Db      1163 sTrgGlyAlaValCysProHlsleuProProCysAlaSerHis----- 1177
QY      251 CTCCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 192
Db      1178 -GlySerTrpLeuSerGlyAlatrgPlyProleuGly---HisArgGlyArgHlsleuG 1196
QY      191 CATG-----GCTTCACAGACATCCACAGCTCAGCCAG 158
Db      1196 yLeuGlyThrGlyTyrArgAspSerGlyGlyLeuAspGlnHlsEserSeValAlaArg 1215

RESULT 15
Q9V9W8 PRELIMINARY; PRT; 815 AA.
AC Q9V9W8;
DT 01-MAY-2000 (TrEMBLrel. 13, Created)
DT 01-MAY-2000 (TrEMBLrel. 13, Last sequence update)
DT 01-JUN-2002 (TrEMBLrel. 21, Last annotation update)
DE CG11518 protein (LD18280P) (Pygopus).
OS Drosophila melanogaster (Fruit fly).
OC Arthropoda; Insecta; Diptera; Brachycera; Muscomorpha;
OC Ephydroidea; Drosophilidae; Drosophila.
OX NCBI_TaxID=7227;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=BERKELEY;
RC MEDLINE=20196006; PubMed=10731132;
RA Adams M.D., Celniker S.E., Holt R.A., Evans C.A., Gocayne J.D.,
RA Amanalides P.G., Scherer S.E., Li P.W., Hoskins R.A., Galie R.F.,
RA George R.A., Lewis S.E., Richards S., Ashburner M., Henderson S.N.,
RA Sutcliffe G., Wortman J.R., Yandell M.D., Zhang Q., Chen L.X.,
RA Brandon R.C., Rogers J.-H.C., Blazer R.G., Champe M., Peiffer B.D.,
RA Man K.H., Doyle C., Baxter E.G., Helt G., Nelson C.R., Miklos G.L.G.,
RA Abril J.F., Agbayani A., An H.-J., Andrews-Fiankoch C., Baldwin D.,
RA Ballev R.M., Basu A., Baxendale J., Bayraktaroglu L., Beasley E.M.,
RA Beeson K.Y., Benos P.V., Berman B.P., Bhandari D., Bolshakov S.,
RA Bokova D., Botchan M.R., Bouck J., Brokstein P., Brothier P.,
RA Butts K.C., Busam D.A., Butler H., Cadieu E., Cantler A., Chandra I.,
RA Chertov J.M., Cawley S., Dahlke C., Davenport L.B., Davies P.,
RA Glodok A., Gong F., Gorrell J.H., Gu Z., Guan P., Harris M.,
RA Harris N.L., Harvey D., Heiman T.J., Hernandez J.R., Houck J.,
RA Hostin D., Houston K.A., Howland T.J., Wei M.-H., Ibbegam C.,
RA Jatali M., Kalush F., Karpen G.H., Ke Z., Kemison U.A., Ketchum K.A.,
RA Kimmel B.E., Kodira C.D., Kraft C., Kravitz S., Kulp D., Lai X.,
RA Laoko B., Lei Y., Levitsky A.A., Li J., Li Z., Liang Y., Lin Z.,
RA Liu X., Mattei B., McIntosh T.C., McLeod M.P., McPherson D.,
RA Merkulov G., Milehina N.V., Mobarry C., Morris J., Moshrefi A.,

```

```

RA Mount S.M., Moy M., Murphy B., Murphy L., Muzny D.M., Nelson D.L.,
RA Nelson D.R., Nelson K.A., Nixon K., Nusser D.R., Pacleb J.M.,
RA Palazzolo M., Pittman G.S., Pan S., Pollard J., Puri V., Reese M.G.,
RA Reinert K., Remington K., Saunders R.D.C., Scheller F., Shen H.,
RA Shue B.C., Siden-Klamas I., Simpson M., Skupski M.P., Smith T.,
RA Spier E., Spirdling A.C., Stapleton M., Strong R., Sun E.,
RA Svitek R., Teocor C., Turner R., Venter E., Wang A.H., Wang X.,
RA Wang Z.-Y., Wasserman D.A., Weinstock G.M., Weissenbach J.,
RA Williams S.M., Woodage T., Worley K.C., Wu D., Yang S., Yao Q.A.,
RA Ye J., Yeh R.-P., Zaveri J.S., Zhan M., Zhang G., Zhao Q., Zheng L.,
RA Zheng X.H., Zhong F.N., Zhong W., Zhou X., Zhu S., Zhu H.O.,
RA Gibbs R.A., Myers E.W., Rubin G.M., Venter J.C.;
RT "The genome sequence of Drosophila melanogaster.";
RL Science 287:2185-2195 (2000).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=BERKELEY;
RA Stapleton M., Brokstein P., Hong L., Agbayani A., Carlson J.,
RA Champe M., Chavez C., Dorsett V., Farfan D., Frise E., George R.,
RA Gonzalez M., Guarin H., Li P., Liao G., Miranda A., Mungall C.J.,
RA Nunoo J., Pacleb J., Paragas V., Park S., Phouanavong S., Wan K.,
RA Yu C., Lewis S.E., Rubin G.M., Celniker S.,
RL Submitted (OCT-2001) to the EMBL/Genbank/DBJ databases.
RN [3]
RP SEQUENCE FROM N.A.
RC TISSUE=HEAD;
RA Cardigan K.M., Parker D.S., Jemison J., Klinedienst S., Kohen R.,
RT "The Role of the Gammy Legs Gene in Wingless Signaling.";
RL (in) Unknown A. (eds.);
RL 42nd Annual Drosophila Research Conference, pp.144-150,
RL Unknown Publisher, Unknown city (2001).
RN [4]
RP SEQUENCE FROM N.A.
RC TISSUE=HEAD;
RA Parker D.S., Jemison J., Cardigan K.M.;
RL Submitted (JAN-2002) to the EMBL/Genbank/DBJ databases.
DR EMBL; AE003778; AAF57161.1;
DR EMBL; AY058500; AAL13729.1;
DR EMBL; AY075095; AAL79357.1;
DR Flybase; Fgn0028951; BEST.LD21971.
DR InterPro; IPR002965; P_tich_extensn.
DR InterPro; IPR001412; tRNA-synt_1.
DR Pfam; PF006628; PHD; 1.
DR PRINTS; PR01217; PRICEXTENSN.
DR SMART; SM00249; PHD; 1.
DR PROSITE; PS00178; AA_TRNA_LIGASE_1; UNKNOWN 2.
DR SEQUENCE 815 AA; 80493 MW; 369PDS5A5D34BC136 CRC64;
SQ

Alignment Scores:
Pred. No.: 0.0276 Length: 815
Score: 120.00 Matches: 108
Percent Similarity: 30.748 Conservative: 34
Best Local Similarity: 23.388 Mismatches: 136
Query Match: 4.69% Indels: 155
DB: Gaps: 20

US-09-989-919-15 (1-1397) x Q9V9W8 (1-815)
QY 1294 CCAACAGAGTACCAAGATTTGATATTCAACTGCAGAGAGGGGAGAAAGCCGTGTT 1235
Db 39 ProGlyAlaArgArgysThrSerSeralAlaAAsnSer-----AlaAlaAlaVal 55
QY 1234 TTGGCTGCTTGAAGGCTTCATTTGAAGCAAGTACAGCAAGAGAGAGAGAGAGAGAG 1175
Db 56 AlaAlaAlaAlaAlaAlaAlaAlaAlaAlaAlaAlaAlaAlaAlaAlaAlaAlaAla 75
QY 1174 CGCCGAGGCTGCTGCTCAACACAGCTGCTCCAGCCAGCTGCTGCTGCTGCTGCTGCTG 1115
Db 76 ThrProGlnAsp-----LeuLeuProProProPrometGlyGlyPheGly 90
QY 1114 GGTGTGACCTGTCGCCCTGAGCCCTG----- 1088

```

Db 91 AspThrIleIleAlaSerAsnProPheAspSerProGlnValSerAlaMetSerSer 110  
 QY 1088 ----- 1088  
 Db 111 SerAlaAlaAlaMetAlaAlaMetAsnGlnMetGlyGlyProGlyGlyGlnHis 130  
 QY 1087 -----GGTGGCCTGGAGAGTCCCTG-----TCGGGA 1059  
 Db 131 PheGlyGlyGlyProGlyGlnHisProHisIleTropLysPargMetGlyMetGlyGly 150  
 QY 1058 GGTCAAGATAGTCCCTCCAGTACAGAAATCACCCATCTCTGGAGATCCCTCCCAAG 999  
 Db 151 GlyProProProProHisMetHisProHisMetHisProHisHisPro----- 167  
 QY 998 TCTCTGTCTCAGACTTAAGTACAGAACTCAGATGATTTGCTCCACGACGAGCC 939  
 Db 168 -----GlyGlyProMetGlyHisProHisGlyProHis 178  
 QY 938 CACCGCTGCTGTGGCATGTGACAGACAGACAGAGGAGGCTTTAGCTACATTCACACT 879  
 Db 178 sProHisMetGlyGly-----ProProProMetArgGlyMetSer----- 191  
 QY 878 ATTGAACACAGTATGGGGCTGCCACAGAACCGTCCAGAGAGGGACACAGCCCCAGT 819  
 Db 192 -----ProMetHisProHisGlnMetGlyProGlyProGlyVa 204  
 QY 818 GGGGAGACTAGCTGAGCTTGTACAGCCACAGAGGTTGAGAGGAGT----- 769  
 Db 204 IGIyleuPro-ProHisMetAsnHisGlyArgProGlyGlyProGlyGlyProGlyGlyP 224  
 QY 768 -----CATACCCCATG-----ATTCTGAGTACAGAGCCAGAGTCCCATAG 726  
 Db 224 roValProMetGlySerProMetGlyGlyIleAlaGlyMetGlyGlyMetSerProMetG 244  
 QY 725 CTAGGCTGGTAACTACATTTTAACATTGGCATTTGACAGTTTGTCTGTGTCACAT 666  
 Db 244 Iy-Gly-----MetGlyGlyProSer-IleSerProHisHisMe 256  
 QY 665 CTGTCTGTGGAGATTTAGCTTTGTCTGAGGGGT-----CCGAGTGACA 618  
 Db 256 tGIyleMetGlyGlyIleuSerProMetGlyGlyProAsnGlyProAsnProArg----- 274  
 QY 617 TGCAGTCTCCAGTCAAGTCACTCCCTCCACAGAGCCAGCTGTGAGGATATGTC 558  
 Db 275 -----AlaMetGlnGlySerProMetGlyGly----- 283  
 QY 557 AGCTGTACAGCCAGCTTCAGATAGTACAGCAGAGGCTTCACATATGTGCTTTGTT 498  
 Db 284 -ProGlyGlnAsnSerProMetAsnSerLeuProMetGly-SerProMetGlyAsn----- 301  
 QY 497 GGGATCAGGAGCACTTGGCTCCATCCAGTCTTGAAAAAGACCAAGCTGCCAGCTGGGG 438  
 Db 302 -----ProIleGlySer-----ProLeuGlyProProSerG 312  
 QY 437 GCAGAGGTGAGAGTACAGAGGCTGATGTTTCTTTCTGGAGGCAAGACATTCCTCC 378  
 Db 312 IyProGlyPro-GlyAsnProGlyAsnHisGlyGlyProGlnGlnGlnGlnGlnPro 331  
 QY 377 TAAATTGACTGCTTGAAGACAGTATACCTCTCTCATCAGCTCAGACCCCTCTGA 318  
 Db 332 -----ProGlnProProMet----- 336  
 QY 317 CTTCGATGGGAGTACTATCTATCACTCCAGAGAGGCTCCGCTCTCTTGAA 258  
 Db 337 -----AsnAsnGlyGlnMetGlyProProProLeuHis 347  
 QY 257 GGGCTTCCCACTGCTACCTGTCCAGGTGTCTGTGTGACACCAAGAGCCGAC 198  
 Db 348 Ser-----ProLeuGlyAsn-----GlyPro 354  
 QY 197 ACTGGGATGGCTCTCACAGA-----CATCCACAGCTCAGCCAGGTTGAACACTGAA 144  
 Db 355 ThrGlyHisGlySerHisMetProGlyGlyProIleProGlyProGlyProGly 374

QY 143 GAGGAGCTCCCGTGTGTAGATGGCCAGAGCGGTAGTCTTCTGGGGA 90  
 Db 375 GlyLeuValGlyProGlyGlyIleSerProAlaHisGlyAsnAsnProGlyGly 392

Search completed: June 24, 2003, 13:57:50  
 Job time: 155 secs





DR HSSP: P19491; 1GR2.  
 DR MGD; MGI:95822; Grin2c.  
 DR InterPro; IPR001320; Ion\_glu\_receptor.  
 DR InterPro; IPR001311; SBP/glu\_receptor.  
 DR Pfam; PF00060; lig\_chan; 1.  
 DR ProDom; PD000500; Ion\_glu\_receptor; 1.  
 DR SMART; SM00079; pabe; 1.  
 DR Receptor; Signal; Transmembrane; Postsynaptic membrane; Calcium;  
 KW Ionic channel; Magnesium.  
 FT CHAIN 1 19  
 FT SIGNAL 20 1239  
 FT DOMAIN 20- 553  
 FT TRANSMEM 554 574  
 FT TRANSMEM 597 617  
 FT TRANSMEM 627 647  
 FT TRANSMEM 815 835  
 FT SITE 612 612  
 FT CARBOHYD 70 70  
 FT CARBOHYD 337 337  
 FT CARBOHYD 438 438  
 FT CARBOHYD 539 539  
 SQ SEQUENCE 1239 AA; 135420 MW; 793E8E731E20C3C9 CRC64;

## Alignment Scores:

Pred. No.: 0.037 Length: 1239  
 Score: 126.00 Matches: 94  
 Percent Similarity: 30.63% Conservative: 23  
 Best Local Similarity: 24.61% Mismatches: 112  
 Query Match: 4.92% Indels: 153  
 Gaps: 24

US-09-989-919-15 (1-1397) x|NME3\_MOUSE (1-1239)

QY 1196 AGAAGAGCCCGGCGCCCGCCAGGCTCGTGGCTCAAAACACACCTGCTCCCTGA 1137  
 DB 927 ArgArgAlaProAlaPro-----ThrThrSerGlyPro----- 937  
 QY 1136 CCCCAGCTTGGCTCATGCTGAGGTGTCACCTCTGCCCTGACCCCTGGG----- 1086  
 DB 938 -----ArgSerCysThrProGlyPro-----ProGlyGlnProSer 949  
 QY 1085 -----CTGGCCTGGAGGTGCTCCCTGTCTGGGAGTGAAGACCTCCCGAGTACGAAT 1029  
 DB 950 ProSerGlyTyrPargProPro-----GlyGlyGlyValArg---ThrProLeuAlaArgArg 966  
 QY 1028 CACCCACATCCTGGAGCATCCCGCCGCAAG-----TCTCCTGTCCAGACTTAAGCT 978  
 DB 967 AlaProGlnProProAlaArgProAlaThrCysAlaGlySerProGlnProAspValSer 986  
 QY 977 GAGAGAACACTCCCGATGATGTCCCGACGACCCCGCCGCTGCTGTGCATGTG 918  
 DB 987 ArgAlaSerCysArgHisAlaTyrPaspAlaArg-----TyrProVal 1000  
 QY 917 CAGACAGACAGACGAGGGGCTTACTTACATTTCTACACTATTGAACACAGTGAAGGCT 858  
 DB 1001 ArgValAlaGlyHisGlyGly-----SerHisLeuSerAlaSer 1012  
 QY 857 GCCCAGGAAACCGCCCGACGAGGAGGACAGCCCCCAGTGGGAGACTAGCTGAGCTTG 798  
 DB 1013 GlnArgArgAlaLeuProGlnArg-----SerLeu 1022  
 QY 797 CTTACAGCCAC-----AGGTTGAGAGGACGATC 768  
 DB 1023 LeuHisAlaHisCysHisTyrSerSerPheProArgAlaGlnArgSerGlyArgProPhe 1042  
 QY 767 ATACCCCATGATTCCTGGAGTAGAGCCGACA-----GTCCCATAGCTAGGC 720  
 DB 1043 LeuPro-----LeuPheProGlnProProGlnProAspAspLeuProLeuGly 1059  
 QY 719 TGGTAAACTCATTTTAAACATTTGGCATTTGTCAGCTGTTGCTGCTCATCTGTCT 660

DB 1060 ---ProGlnGlnLeuAlaArgArgGlnAlaLeuLeuArgAlaAlaTyrPala----- 1075  
 QY 659 GTCTGGAGATGTAGCTTGTCTGCAGGGGTTCCAGG---TGACATGAGGTGTCGCACT 603  
 DB 1076 -----ArgGly- ProArgProArgHisAlaSerLeuProse 1087  
 QY 602 GCAGATC-----ACTCCCTCCACAGGACGACGACG---TCTGAGATGATGTC 558  
 DB 1087 rSerValAlaGlnAlaPheThrArgSerHisProLeuProAlaArgCysThrGlyHisAl 1107  
 QY 557 AGCTGTGACCGGACGCTCCAGATGATGTCAGCCAGAGGCTTCACATATGTGCTTTGT 498  
 DB 1107 AcYsaIaCysProCysProGln----- 1114  
 QY 497 GGGATCAGGGACCACTTGCTCCATCCAGTCTTGAAGAACAGCAGCTGCCGA----- 446  
 DB 1115 -----SerArgProSerCysArgHisValAla 1123  
 QY 445 -----CTGGGGGGGAGAGGTGAGAGGTAAAGAGGCTGAGATG 408  
 DB 1123 aglInhrGlnSerLeuArgLeuProSerTyrArgGlnAlaCysValGlnGlyValProAl 1143  
 QY 407 TTTCTTTCTGGAGGCAAGACCATTTCTCCCTAATTGACTGCTTGAAAGACAGTATAC 348  
 DB 1143 aglyValAlaAla-----ThrTyrGlnhr 1151  
 QY 347 TCTCTCTCAT-----CACTCAGACACCCCTCTGACTTCGATGGGGGATCATAT 297  
 DB 1151 oArgGlnHisValCysLeuHisThrHisThrHisLeuProPheCysTyrGlyThrValCys 1171  
 QY 296 CTATCAACCTCCAGAGAGGTCTGCCATCTCCCTTGAAGAGGCTTCTCCATGCTCAC 237  
 DB 1171 sArgHisProProProCysSerSerHisSerProThr----- 1183  
 QY 236 CTGTCCAGGTGCTGTGTTGT-----GACCAAGAGCCCGACACTGGG 192  
 DB 1184 -----LeuIleGlyThrTyrGlnProProSerHisArgGlyArgThrLeuGln 1199  
 QY 191 CANG 188  
 DB 1199 yLeu 1200

RESULT 2  
 POS RUBVR STANDARD: PRT: 1063 AA.  
 AC P19725; Q86370; Q86371; Q86372;  
 DT 01-FEB-1991 (Rel. 17, Created)  
 DT 01-FEB-1991 (Rel. 17, Last sequence update)  
 DT 16-OCT-2001 (Rel. 40, Last annotation update)  
 DE Structural polypeptide [Contains: Nucleocapsid protein C; Membrane glycoprotein E2; Membrane glycoprotein E1].  
 OS Rubella virus (vaccine strain RA27/3).  
 OC Viruses; ssRNA positive-strand viruses, no DNA stage; Togaviridae;  
 OC Rubivirus.  
 OX NCBI\_TaxID=11044;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE=89286505; PubMed=2740235;  
 RA Nakhasi H.L., Thomas D., Zheng D., Liu T.Y.;  
 RT "Nucleocapsid sequence of capsid, E2 and E1 protein genes of Rubella virus vaccine strain RA27/3."  
 RL Nucleic Acids Res. 17:4393-4394(1989).  
 CC - FUNCTION: GLYCOPROTEIN E1 CONTAINS THE VIRAL HEMAGGLUTININ ACTIVITY.  
 CC - SUBUNIT: MULTIPLE COPIES OF THE C PROTEIN COMPRISE THE NUCLEOCAPSID.  
 CC - DISEASE: RUBELLA INFECTION DURING THE FIRST TRIMESTER OF PREGNANCY RESULTS IN DEATH OR MULTISYSTEM BIRTH DEFECTS & MICROCEPHALY.  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way



CC -1- FUNCTION: GLYCOPROTEIN E1 CONTAINS THE VIRAL HEMAGGLUTININ  
 CC ACTIVITY.  
 CC -1- SUBUNIT: MULTIPLE COPIES OF THE C PROTEIN COMPRISE THE  
 CC NUCLEOCAPSID.  
 CC -1- DISEASE: RUBELLA INFECTION DURING THE FIRST TRIMESTER OF PREGNANCY  
 CC RESULTS IN DEATH OR MULTISYSTEM BIRTH DEFECTS & MICROCEPHALY.  
 CC  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC  
 CC EMBL: M30776; AAA47421.1;  
 CC PIR: J00087; GNMV77.  
 DR Polyprotein; Nucleocapsid; Glycoprotein; Transmembrane; Signal.  
 KM CHAIN 1 300 MEMBRANE GLYCOPROTEIN C.  
 FT CHAIN 301 562 MEMBRANE GLYCOPROTEIN E2.  
 FT SIGNAL 563 582  
 FT CHAIN 583 1063  
 FT TRANSMEM 278 300  
 FT TRANSMEM 517 535  
 FT TRANSMEM 538 555  
 FT CARBOHYD 353 353  
 FT CARBOHYD 371 371  
 FT CARBOHYD 429 429  
 FT CARBOHYD 458 458  
 FT CARBOHYD 759 759  
 FT CARBOHYD 791 791  
 SQ SEQUENCE 1063 AA; 114713 MW; 18E7B28B9A667 CRC64;

Alignment Scores:  
 Pred. No.: 0.128 Length: 1063  
 Score: 119.00 Matches: 98  
 Percent Similarity: 32.46% Conservative: 39  
 Best Local Similarity: 23.22% Mismatches: 127  
 Query Match: 4.65% Indels: 159  
 DB: 1 Gaps: 22

US-09-989-919-15 (1-1397) x: POLS\_RUBVH (1-1063)

QY 1184 GCCCCACCCCGCCCGCCAGCTCTGCTCAAC-----ACACTGCTCCCTGACCC 1134  
 DB 77 AlarProProProProGluGluArgGlnGlnSerArgSerGlnThrProAlaProLysPro 96  
 QY 1133 CAGCTTGGCTCAGCTGAGGTGTGCACTCTGCCCTGACCCCTGGGCTGGCTGGAG 1074  
 DB 97 SerArg-----AlaPro-----ProGlnGlnProGln----- 105  
 QY 1073 TGTCCCTGTGGAGGTAGAGATAGCTCCAGGTAG-----AAT 1029  
 DB 106 TGTCCCTGTGGAGGTAGAGATAGCTCCAGGTAG-----AAT 1029  
 QY 1028 CACCCACATCTGTGAGCATCCCCCGCAAGCTCTGTCTCCAGACTTAAGTGAAGA-- 972  
 DB 117 AlarProArgProGluLeuGlyProProThrSerProGlnAlaAlaValAlaAlaArgGly 136  
 QY 971 -----ACACTCCAGATGATTTGCCACGACGACCCCGCCAGCTGCTGT----- 926  
 DB 137 LeuArgProProLeuHisAspPro--AspThrGlu-AlaProThrGlnAlaCysValTh 155  
 QY 925 -----GGCATGTGACGACGACGACGAGGGCTTTAGTCACTTCTCACTATTGAACAC 870  
 DB 155 rSerTrpLeuTrpSerGlnGlyGlnGlyAlaValAlaPheTyrArgValAspLeuHisPheTh 175  
 QY 869 AGTATGGGCTGCTCCCGACGGAACGCTCCAGGGA----- 836  
 DB 175 rAsnLeuGlyThrProProLeuAspGlyArgTrpAspProAlaLeuMetTyrAs 195  
 QY 835 -----GGGACACGACCCCGACGAGTGGGAGACTAGCTGACTTGTCTTACAGCCCAAGG 783

DB 195 nProCysGlyProGlnProPro--AlaHisValValArgAlaTyrAsnGlnProAlaG 214  
 QY 782 GTTGAGAGGAGCATCATCCCCCATGATTCCTGGACATGAGCCAGACGATCCCATAGCTA 723  
 DB 214 yAspValArg-----GlyValTrpGlyValGlnArgGlnTyr-Ala- 228  
 QY 722 GGCTGTAAACTCATATTTAACAT-----TGCACTTATTTGACAGTTGTCTCC 675  
 DB 229 -----GlnGlnAspPheArgValGlyGlyThrArgTrpHisArgLeuLeuMetP 246  
 QY 674 TGTGCATCATCTGTCTGTGGAGATTTAGCTTTGTCTGCAGAGGTTCCAGGTGCATGC 615  
 DB 246 roval-----ArgGlyLeuAspGlyAsp----- 253  
 QY 614 AGTCTGCACATGCAAGTCACTCCCTCCACAGGACGAGCCAGCTGTGAGATG-TCAG 556  
 DB 254 -----ThraAlaProLeuProProHisThrThrGlnArgGlyLeuThrArgSerA 270  
 QY 555 CTGTGCAGCGG--AGCCCTCAGATAGGTTCAGCCAGAGGCTTTCACATATGTGCTTGT 499  
 DB 270 lArgArgProTrpArgGlyLeArgPheGlyAlaProGlnAlaPheLeuAlaGlyLeuLeu 290  
 QY 498 TGGATTCAGGACCACTGTGCTTCATCCAGTCTTGAAAAAGACCAAGCTGCGGAGCTGGGG 439  
 DB 290 euAlaAlaValAla-ValGlyThrAlaArgAlaGlyLeuGlnProArgAlaAspMet--- 308  
 QY 438 GCGAAGGTGAGAGTAACAGAGGCTGATGTTCTCTTCTTGTGAGGCAAGCATTCCTCC 379  
 DB 308 ----- 308  
 QY 378 CTAATTGACGTCTTGAAGACAGATATACCTCTCTCATCATCCTCAGACCCCTCTG 319  
 DB 309 -----AlaAlaProPromet 313  
 QY 318 ACTTCGATGGGGATCACTATCTATCAACTCCAGAGAGCTCTGCACTCTCTTGA 259  
 DB 314 -----Pro-ProGlnPro----- 317  
 QY 258 AGGCTTCCCGACCTGAGCTCAGCTGTCAGAGGATGCTGTGGTGAACCAAGCCCGCA 199  
 DB 318 -----ProArgAlaHisGlyGlnHisTyrGlnHisHisHisGlnLeuProPh 334  
 QY 198 CACTGGCATGAGCTCTCAGACATCCACAGCTCAGCCAGGTGAACCTGAAGAGAG 139  
 DB 334 eleuGlyHisAspGlyHisHis-----GlyG 343  
 QY 138 CAGCTCCGTGTGTGATGATGCCAGCAGCGGTAGT-----CT 100  
 DB 343 yThrLeuArgValGlyGlnHisHisArgGlnAlaSerValLeuProGlyHisTyrLe 363  
 QY 99 TCTGTGGGATGTGTGTCTGTGGA-----TACACTGTGTACTCGTACTGTC 54  
 DB 363 uGlnGlyGlyTrpGlyCysTyrAsnLeuSerAspTrpHisGlnGlyThrHisValCys 382

RESULT 4  
 POLS\_RUBVT STANDARD; PRT; 1063 AA.  
 AC P07566;  
 DT 01-APR-1988 (Rel. 07, Created)  
 DT 01-AUG-1988 (Rel. 08, Last sequence update)  
 DT 16-OCT-2001 (Rel. 40, Last annotation update)  
 DE Structural polyprotein [contains: Nucleocapsid protein C; Membrane glycoprotein E2; Membrane glycoprotein E1].  
 OS Rubella virus (strain Thierien)  
 OC Viruses; ssRNA positive-strand viruses, no DNA stage; Togaviridae;  
 OC Rubivirus.  
 OX NCBI\_TaxID=11045;  
 RN [1].  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE=90281585; PubMed=2353453;  
 RA Dominguez G., Wang C.Y., Frey T.K.;  
 RT "Sequence of the genome RNA of rubella virus: evidence for genetic rearrangement during togavirus evolution.";





QY 365 CTTGAAGACACGATATCTCTCTATCATCTGACAGCCCTCTGACTTCGATGGG 306  
 Db 309 .....  
 QY 305 GATCATATCTATCAACCTCCAGAGAGGTCTCCATCTCCCTTGAAGGGCTTCCCA 246  
 Db 318 .....  
 QY 245 CTGGCTACCTGTCAGGTGCTGTTGTCACCAAGGCCGACACTGGCATGGC 186  
 Db 319 CVALAHIEGLYGLNHYGLNHYGLNHYGLNHYGLNHYGLNHYGLNHYGLNHYGLNHY 338  
 QY 185 TCTCAGACATCCACAGCTCCAGCTTGAACACTGAAGAGGAGCTCCCTGGT 126  
 Db 339 GLYHSHIS-.....-GLYHSHIS-.....  
 QY 125 GTAGATGGCGACAGCGGTAGT-.....-CTTCTGGAGGATGG 87  
 Db 348 GLYGLNHYGLNHYGLNHYGLNHYGLNHYGLNHYGLNHYGLNHYGLNHYGLNHY 367  
 QY 86 TGCTGCTGGGA-.....-TACACTGTACTCGGTACTGC 54  
 Db 368 GLYCYETRYASNSUSERAAPTTPHISGLNGLYTHRHISVALCYS 382  
 RESULT 5  
 CYRB HUMAN STANDARD; PRT; 897 AA.  
 AC P32937;  
 DT 01-OCT-1993 (Rel. 27, Created)  
 DT 01-FEB-1996 (Rel. 33, Last sequence update)  
 DT 16-OCT-2001 (Rel. 40, Last annotation update)  
 DE Cytokine receptor common beta chain precursor (CDW131 antigen).  
 GN CSF2RB OR IL5RB OR IL3RB.  
 OS Homo sapiens (Human).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.  
 OC NCBI\_TaxID=9606;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE=91088571; PubMed=1702217;  
 RA Hayashida K., Kitamura T., Gorman D.M., Arai K., Yokota T.,  
 RA Miyajima A.;  
 RT "Molecular cloning of a second subunit of the receptor for human  
 RT granulocyte-macrophage colony-stimulating factor (GM-CSF):  
 RT reconstitution of a high-affinity GM-CSF receptor.";  
 RL Proc. Natl. Acad. Sci. U.S.A. 87:9655-9659(1990).  
 RN [2]  
 RP REVISION TO 454.  
 RA Kitamura T.;  
 RL Submitted (FEB-1991) to the EMBL/GenBank/DBJ databases.  
 CC - FUNCTION: HIGH AFFINITY RECEPTOR FOR INTERLEUKIN-3, INTERLEUKIN-5  
 CC AND GRANULOCYTE-MACROPHAGE COLONY-STIMULATING FACTOR.  
 CC - SUBUNIT: HETERODIMER OF AN ALPHA AND A BETA CHAIN. THE BETA  
 CC CHAIN IS COMMON TO THE IL-3, IL-5 AND GM-CSF RECEPTORS.  
 CC - SUBCELLULAR LOCATION: Type I membrane protein.  
 CC - SIMILARITY: CONTAINS 2 FIBRONECTIN TYPE III-LIKE DOMAINS.  
 CC - SIMILARITY: BELONGS TO THE CYTOKINE FAMILY OF RECEPTORS.  
 CC - DATABASE: NAME=PRO; NOTE=CD guide CDW131 entry;  
 CC WWW=http://www.ncbi.nlm.nih.gov/prov/cd/cdw131.htm.  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See http://www.isb.ch/announce/  
 CC or send an email to license@isb-sib.ch).  
 CC  
 DR EMBL; M59941; AAA18171.1;  
 DR PIR; A39255; A39255;  
 DR HSSP; P19235; IBA;  
 DR Genew; HGNC; 2436; CSF2RB.

DR MIM; 138981;  
 DR InterPro; IPR002966; CRIA.  
 DR InterPro; IPR000282; Cytok\_receptor\_2.  
 DR InterPro; IPR003961; FN\_III.  
 DR InterPro; IPR003531; Hemtopoptn\_S\_F1.  
 DR Pfam; PF00041; fn3; 2.  
 DR SMART; SMO0060; FN3; 2.  
 DR PROSITE; PS01355; HEMATOPO\_REC\_S\_F1; 1.  
 DR Receptor; Transmembrane; Glycoprotein; Repeat; Signal.  
 FT SIGNAL 1 16  
 FT CHAIN 1 897  
 FT DOMAIN 17 443  
 FT TRANSMEM 444 460  
 FT DOMAIN 461 897  
 FT DOMAIN 129 238  
 FT DOMAIN 336 434  
 FT DISULFID 35 45  
 FT DISULFID 75 91  
 FT CARBOHYD 58 58  
 FT CARBOHYD 191 191  
 FT CARBOHYD 346 346  
 SQ SEQUENCE 897 AA; 97335 MW; 3398E37FDB8F93A CRC64;

Alignment Scores:  
 Pred. No.: 0.15 Length: 897  
 Score: 118.00 Matches: 104  
 Percent Similarity: 34.37% Conservative: 51  
 Best Local Similarity: 23.06% Mismatches: 152  
 Query Match: 4.61% Indels: 144  
 DB: 1 Gaps: 22

US-09-989-919-15 (1-1397) x CYRB\_HUMAN (1-897)  
 QY 1221 GCGTTTCATTGAAGCAATGAGCGAAGAGCCCGCCGACCCCGCCGAGTCTCT 1162  
 Db 464 GYLIETRYGLTYARGLNHYGLNHYGLNHYGLNHYGLNHYGLNHYGLNHYGLNHY 483  
 QY 1161 GGC-.....-TCAACACACCGCTGCTCCGACCCGACGTCGTGCTCATGTCGAGTG 1111  
 Db 484 SerHisLeuPheGlnHisHisHisHisHisHisHisHisHisHisHisHisHisHis 503  
 QY 1110 TGCACCTCTGCCCCCT-.....-GACCCCTGGGCTGGCTG-.....-GGA 1075  
 Db 504 ThrSerGlySerProProHisGlnGlyProTyrGlySerArgPheProGluLeuGlnGly 523  
 QY 1074 GGTGCTCCCTGTC-.....-GGAGGTCAAGATAGCTCCCGACAGTACAGATACCCACA 1021  
 Db 524 ValPheProValGlyPheGlyAspSerGluVal-.....-SerProLeu 537  
 QY 1020 TCGTGAAGATCCCGCCGACGATCTCTGTCAGACCTAAGCTGAGAGAACCTCCACGA 961  
 Db 538 ThrIleGluAspProHisHisHisHisHisHisHisHisHisHisHisHisHisHis 557  
 QY 960 TGGATT-.....-GTCCCCACG-.....-CAGCCCCCAGCGCTGCTGGCA 922  
 Db 558 AlaSerAspLeuProHisGlnGlnProProSerProGlnProGlyProProAlaAlaSer 577  
 QY 921 TGTGACGACGACGACGACGAGGGCTTAGCTA-..... 891  
 Db 578 HisThr-ProGluGlyGlnAlaSerSerPheAspPheAsnGlyProTyrLeuGlyPro 597  
 QY 890 -CATTCACACATATGAACACAGATAGGGGCTGCGCCAGAGAGCGTCCCGAGGGGG 832  
 Db 597 ohHisSerArgSer-LeuProAspIleLeuGlnGlnProGluProProGlnGlnGlyGly 617  
 QY 831 CACAG-.....-CCCGCAGTGGGAGAGCTAGCTGAGCTTGTATACAGCCACAGGGCTGG 778  
 Db 617 erGlnHisSerProProProGlySerLeuGlnGlyTyrLeuGlySerLeu-ProAlaGlyGly 636  
 QY 777 AGAGGAGTCAATACCCCATGATTCCTGAGCTAGAGCAGCAGCATGCTCATGCTAGGCTG 718  
 Db 636 InvalGlnLeuValPro-..... 641

QY 717 GTAACCTACATTAACTTTGACATTGACAGCTTTGCTCTGTCACATCTGCTGT 658  
 Db 641 ----- 641  
 QY 657 CTGGAGTGTAGCTTTCTGACAGGCTTCCAGTGCATGACAGTCCAGTGCAG 598  
 Db 642 -----LeuAlaGlnAlaMetGlyProGlyGln-----AlaValGlnValGln 656  
 QY 597 TCACTCCCTCCACAGCAAGCCAGCTGCTGAGATGATGAGTGTGCTGACGAGCCCTC 598  
 Db 656 rgrgrProSerGingly-----AlaAlaGlySerProSerLeug 669  
 QY 537 AGATAGCTGACGAGAGCCCTTACATATGTGCTTTGTTGGATGACGAGCATTGAC 478  
 Db 669 luserGlyGly-----GlyProAlaProProAla 679  
 QY 477 TCCATCCAGCTTGAAGAAAGACAGCTGCCGCTGGGGGGGAGAGTGAAGTACAG 418  
 Db 679 euGlyPro-----ArgValGlyGlyGlnAspGlnLysAspSer 692  
 QY 417 GGCTGATGTTCTCTTCTTGAGGCAAGACATTCCTCAATTTGATGCTGCTGAGA 358  
 Db 692 roValAlaIleProMetSerSerGlyAspHrGlnAspProGlyValAlaSerGlyTyv 712  
 QY 357 CACGATACCTCTCTCTCATCTGACCTGACACCCCTCTGACTCCGATGGGGAGTACTA 298  
 Db 712 alser-SerAlaAspLeuVal-----PheHrProAsnSerGlyAlaSer 726  
 QY 297 TCTATCAACTC-----CAGAGAGTCTCTGCCATCTCCCTT 262  
 Db 727 SerValSerLeuValProSerLeuglyLeuProSerAspGlnHrProSerLeuCySpro 746  
 QY 261 GGAAGGCTTCT-----CCCACTGGCTACCTGCTCAGGTG----- 226  
 Db 747 GlyLeuAlaSerGlyProProGlyAlaProGlyProValLysSerGlyPheGlyGlyTyv 766  
 QY 225 GTCTGTTGTGACCAAGAGCCGACACTGGGAGTGTCTTCACAGATCCACAGCC 166  
 Db 767 ValGlnLeuProProIleGlnGlyArg-SerProArgSerProArgAsnAsnProValPr 786  
 QY 165 T-----CAGCAGGTTGAACACTGAAGAGGAGGAGCTCCG 130  
 Db 786 oProGlnAlaLysSerProValLeuAsnProGly-----Gluar 799  
 QY 129 TGGTGTGAGATGCCAGCAGCGGATGCTTCTCGGGGGAGTGTGCTGTGGATACAC 70  
 Db 799 gProAlaAspValSerProThiSerProGlnProGlyLeuLeuValLeuGlnGlnva 819  
 QY 69 TGGTACTCGTACTGCTGCTGCG 45  
 Db 819 lGlyAap---TyrCySpheLeuPro 826

RESULT 6  
 ICP0\_HSV11  
 ID ICP0\_HSV11 STANDARD; PRT; 775 AA.  
 AC P08393;  
 DT 01-AUG-1988 (Rel. 08, Created)  
 DT 01-AUG-1988 (Rel. 08, Last sequence update)  
 DT 15-JUN-2002 (Rel. 41, Last annotation update)  
 DE Trans-acting transcriptional protein ICP0 (Immediate-early protein IE110) (VNM110) (Alpha-0 protein).  
 GN ICP0 OR IE110.  
 OS Herpes simplex virus (type 1 / strain 17).  
 OC Viruses; dsDNA viruses, no RNA stage; Herpesviridae;  
 OC Alphaherpesvirinae; Simplexvirus.  
 NC NCBL\_TaxID=10299;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE=88274327; PubMed=2839594;  
 RA McGeoch D.J., Dalrymple M.A., Davison A.J., Dolan A., Frame M.C., McNab D., Perry L.J., Scott J.E., Taylor P.;  
 RT "The complete DNA sequence of the long unique region in the genome of herpes simplex virus type 1.";

RL J. Gen. Virol. 69:1531-1574(1988).  
 RN [2]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE=87059760; PubMed=3023529;  
 RA Perry L.J., Rixon F.J., Everett R.D., Frame M.C., McGeoch D.J.;  
 RT "Characterization of the IE110 gene of herpes simplex virus type 1.";  
 RL J. Gen. Virol. 67:2365-2380(1986).  
 RN [3]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE=89036163; PubMed=2846760;  
 RA Perry L.J., McGeoch D.J.;  
 RT "The DNA sequences of the long repeat region and adjoining parts of the long unique region in the genome of herpes simplex virus type 1.";  
 RL J. Gen. Virol. 69:2831-2846(1988).  
 CC -1- FUNCTION: ICP0 IS A GENERAL TRANSACTIVATOR OF ALL THREE CLASSES OF HSV GENES WHICH FUNCTIONS SYNERGISTICALLY WITH ICP4. OFTEN TERMED AS A PROMISCUOUS TRANSACTIVATOR. MAY PLAY A ROLE IN THE BALANCE BETWEEN THE LATENT AND LYtic STATES, REACTIVATING LATENT HSV. ICP0 MIGRATES TO THE ND10 NUCLEAR BODIES IN THE EARLY TIMES OF INFECTION AND DISRUPTS THEM. BINDS TO HAUSP AND MAY MODIFY ITS DEUBIUTINATING SUBSTRATE SPECIFICITY OR ACTIVITY ON VIRAL OR CELLULAR TARGETS, LEADING TO AN INCREASED OR DECREASED STABILITY OF THESE PROTEINS.  
 CC -1- SIMILARITY: CONTAINS 1 RING-TYPE ZINC FINGER.  
 CC -1- SIMILARITY: TO OTHER HERPESVIRUSES ICP0 PROTEIN.  
 CC CC  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC CC  
 DR EMBL; X14112; CAA32336.1; -  
 DR EMBL; X14112; CAA32293.1; -  
 DR EMBL; X04614; CAA28265.1; -  
 DR PIR; A29152; EDBE11.  
 DR InterPro; IPR001841; Znf\_Ring.  
 DR Pfam; PF00097; zf-CHC4; 1.  
 DR SMART; SM00184; RING; 1.  
 DR PROSITE; PS00518; ZF\_RING\_1; 1.  
 DR PROSITE; PS50089; ZF\_RING\_2; 1.  
 KW Transcription regulation; Trans-acting factor; Activator; Zinc-finger;  
 KW DNA-binding; Early protein.  
 FT ZN FING 116 157 RING-TYPE.  
 FT DOMAIN 233 243 POLY-ASP.  
 FT DOMAIN 305 308 POLY-GLY.  
 FT DOMAIN 558 568 POLY-SER.  
 SQ SEQUENCE 775 AA; 78456 MW; DF38A1C539DAB15C CRC64;

Alignment Scores:  
 Pred. No.: 0.176 Length: 775  
 Score: 117.00 Matches: 47  
 Percent Similarity: 36.00% Conservative: 7  
 Best local Similarity: 31.33% Mismatches: 61  
 Query Match: 4.57% Indels: 35  
 DB: 1 Gaps: 8

US-09-989-919-15 (1-1397) x ICP0\_HSV11 (1-775)  
 QY 1188 CCGGCGCCACCCCGCCAGTCTGCTCAAAACACACCTGCTCCCTGACCCGAGTC 1129  
 Db 381 ProGlySerAlaProArgProGlyProProAlaSerAlaAlaSer---GlyProAla 399  
 QY 1128 TTGGCTCATGCTGAGGTG-----TGACCTCTGCGCTGACCCCTGGGCTGGCTGGCA 1075  
 Db 400 ArgProArgAlaAlaValAlaProGlySerAlaGlnAlaProPro----- 414  
 QY 1074 GTGTCCCTGTGGGGAGTCAAGATAGCTCCCGAGTACAGATCAACCATCCTGG 1015  
 Db 415 -----GlyProGlyProArgAlaProAla-ProG 424

CC 1014 AGCATCCCCCGCAAGTCTCTCTCCAGACTTAAGCTGAGAGAACACTCCAGCATGATT 955  
 CC |||||  
 Db 424 yalagluiprolalalargProalaasp-----AlargArgVal----- 437  
 CC |||||  
 Oy 954 GTCCCGACGAGGCCCGCCCGCTGCTGGCATGTGACGACGACGAGGCGCTTTA 895  
 CC |||||  
 Db 438 ----ProGlnSer---HisSerSerLeuAlaGlnAlaAlaasnGlnGlnSerLeu 455  
 CC |||||  
 Oy 894 GCTACATCTTCACACTTATTAACACAGTATGCG-----GCTCCCGACGG 850  
 CC |||||  
 Db 455 sarGAlaArgAlaThrValAlaArgGlySerGlyGlyProGlyValGlnGlyHisGln 475  
 CC |||||  
 Oy 849 AACCGTCCCGACGAGGCGACACGCCCGCTGGGAGACTAGCTGAGCTTGGTTACAG 790  
 CC |||||  
 Db 475 y-----ProSerArgGlyAlaAlaProSerGlyAlaAlaProLeuProSerAlaAla 493  
 CC |||||  
 Oy 789 CCACAGGGTGTGAGAGCGAGCTCATACCC 762  
 CC |||||  
 Db 493 rValGlnGlnGlnAlaAlaValAlaArgPro 502  
 CC |||||  
 RESULT 7  
 DRPL HUMAN STANDARD; PRT; 1185 AA.  
 ID DRPL HUMAN STANDARD; PRT; 1185 AA.  
 AC p54259;  
 DT 01-OCT-1996 (Rel. 34, Created)  
 DT 01-OCT-1996 (Rel. 34, Last sequence update)  
 DT 16-OCT-2001 (Rel. 40, Last annotation update)  
 DE Atrophin-1 (Dentatorubral-pallidoluysian atrophy protein).  
 GN DRPLA.  
 OS Homo sapiens (human).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.  
 OX NCBI\_TaxID=9606;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC TISSUE=Cerebellum, and Brain.  
 RX MEDLINE=95141175; PubMed=7842016;  
 RA Nagatuchi S., Yanagisawa H., Ohsaki E., Shitayama T., Tadokoro K.,  
 RT Inoue T., Yamada M.;  
 RT "Structure and expression of the gene responsible for the triplet  
 RT repeat disorder, dentatorubral and pallidoluysian atrophy (DRPLA).";  
 RL Nat. Genet. 8:177-182(1994).  
 RN [2]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE=96262314; PubMed=8965642;  
 RA Margolis R.L., Li S.-H., Young W.S., Wagster M.V., Stine O.C.,  
 RA Kidwai A.S., Ashworth R.G., Ross C.A.;  
 RT "DRPLA gene (atrophin-1) sequence and mRNA expression in human  
 RT brain.";  
 RL Brain Res. Mol. Brain Res. 36:219-226(1996).  
 RN [3]  
 RP SEQUENCE OF 470-725 FROM N.A.  
 RC TISSUE=Brain cortex;  
 RX MEDLINE=93315145; PubMed=8325628;  
 RA Li S.-H., McInnis M.G., Margolis R.L., Antonarakis S.E., Ross C.A.;  
 RT "Novel triplet repeat containing genes in human Brain: cloning,  
 RT expression, and length polymorphisms.";  
 RL Genomics 16:572-579(1993).  
 CC -1- TISSUE SPECIFICITY: THE LEVELS ARE RELATIVELY HIGH IN THE BRAIN,  
 CC OVARY, TESTIS AND PROSTATE. LOWER LEVELS ARE DETECTED IN THE  
 CC LIVER, THYMUS AND LEUKOCYTES.  
 CC -1- POLYMORPHISM: THE POLY-GLN REGION OF DRPLA IS HIGHLY POLYMORPHIC  
 CC (7 TO 23 REPEATS) IN THE NORMAL POPULATION AND IS EXPANDED TO  
 CC ABOUT 49-75 REPEATS IN DRPLA PATIENTS. LONGER EXPANSIONS RESULT IN  
 CC EARLIER ONSET AND MORE SEVERE CLINICAL MANIFESTATIONS OF THE  
 CC DISEASE.  
 CC -1- DISEASE: DEFECTS IN DRPLA ARE THE CAUSE OF DENTATORUBRAL-  
 CC PALLIDOLUYSIAN ATROPHY, AN AUTOSOMAL DOMINANT NEURODEGENERATIVE  
 CC DISORDER CHARACTERIZED BY A LOSS OF NEURONS IN THE DENTATE  
 CC NUCLEUS, RUBRUM, GLOBUS PALLIDUS AND LUY'S BODY. CLINICAL FEATURES  
 CC ARE MYOCLONIC EPILEPSY, DEMENTIA, AND CEREBELLAR ATAXIA. ONSET OF  
 CC THE DISEASE OCCURS USUALLY IN THE SECOND DECADE OF LIFE AND DEATH

CC IN THE FOURTH.  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (see <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----  
 CC EMBL: D31840; BAA06626.1;  
 CC DR EMBL: U23851; AAB50276.1;  
 CC DR EMBL: L10377; -; NOT ANNOTATED\_CDS.  
 CC DR Genew: HGNC:3033; DRPLA.  
 CC DR MIM: 125370;  
 CC DR InterPro: IPR002951; Atrophin.  
 CC DR Pfam: PF03154; Atrophin-1; 2.  
 CC DR PRINTS: PR01222; ATROPHIN.  
 CC KW Triplet repeat expansion; Polymorphism.  
 CC FT DOMAIN 73 82  
 CC FT POLY-PRO. SER/GLU-RICH (MIXED CHARGE).  
 CC FT 302 305  
 CC FT POLY-PRO.  
 CC FT 376 382  
 CC FT POLY-SER.  
 CC FT 386 397  
 CC FT POLY-SER.  
 CC FT 442 447  
 CC FT POLY-PRO.  
 CC FT 479 483  
 CC FT POLY-HIS.  
 CC FT 484 497  
 CC FT POLY-GLN.  
 CC FT 504 507  
 CC FT POLY-PRO.  
 CC FT 564 574  
 CC FT POLY-SER.  
 CC FT 704 707  
 CC FT POLY-PRO.  
 CC FT 802 815  
 CC FT ARG/ALA-RICH (MIXED CHARGE).  
 CC FT 827 837  
 CC FT ARG/GLU-RICH (MIXED CHARGE).  
 CC FT 925 934  
 CC FT MISSING (IN REF. 2).  
 CC FT 94 94  
 CC FT Y -> H (IN REF. 2).  
 CC FT 333 333  
 CC FT M -> I (IN REF. 2).  
 CC FT 339 339  
 CC FT P -> T (IN REF. 3).  
 CC FT 541 541  
 CC FT G -> A (IN REF. 2).  
 CC FT 1028 1028  
 CC SQ SEQUENCE 1185 AA; 124785 MW; 56C306267331C005 CRC64;  
 CC -----  
 CC Alignment Scores:  
 CC Pred. No. 0.185  
 CC Score: 117.00  
 CC Percent Similarity: 32.23%  
 CC Best Local Similarity: 23.53%  
 CC Query Match: 4.57%  
 CC DB: 1  
 CC Gaps: 20  
 CC -----  
 CC US-09-989-919-15 (1-1397) x DRPL\_HUMAN (1-1185)  
 CC Oy 1190 GCCCGCGCCCGACCC----- 1176  
 CC |||||  
 Db 366 AlaProAlaProProMetArgPheProTyrSerSerSerSerSerAlaAlaAla 385  
 CC -----  
 CC Oy 1175 -----CCGCCAGGTCCTGCTCAACACCA 1149  
 CC |||||  
 Db 386 SerSerSerSerSerSerSerSerSerSerSerAlaSerProPheProAlaSerGlnAlaLeu 405  
 CC -----  
 CC Oy 1148 CCGTCT-----CCCGACCCAGCTTGGCTTAGTGAGGTGACCT 1104  
 CC |||||  
 Db 406 ProSerTyrProHisSerPheProProThrSerLeuSerValSerAsnGlnProPro 425  
 CC -----  
 CC Oy 1103 CTGCCCTACCCCTGGCGCTGGCGGTGTCCTCGGAGGTGACAGATGACCT 1044  
 CC |||||  
 Db 426 LysTyrThrGlnProSerLeuProSerGlnAlaValTyrPheGlnGlyProProProPro 445  
 CC -----  
 CC Oy 1043 CCCGAG-----GTACAGAAATCAACCCATCTGGAGACATCCCGCGCAAG 999  
 CC |||||  
 Db 446 ProProTyrGlyArgIleuLeuAlaAsnSerAlaHisPheGlyProPheProPro----- 464  
 CC -----  
 CC Oy 998 TCTCTGTGCAAGCTTAAGTGAAGACACTCCACAGATGATGCTCCCGACGACCC 939  
 CC |||||  
 Db 465 -----SerThrGlyAlaGlnSerThrAlaHisProProValSerThrHisHis 481



```

Db      448  SerrTrpAlaCysLeuProMetGlyProSerGlnGlnAlaThrCysGlnValThrGlnGln 467
      1025  CCA---CATCTCTGA----- 1014
      468  ProSerHisProGlyProLeuSerGlySerProAlaGlnSerAlaProThrLeuAlaCys 487
      1013  -----GCA 1011
      488  ThrGlnValProLeuValLeuAlaAspAsnProAlaTyrArgSerPheSerAspCysCys 507
      1010  TCCCGCGGCAAGTCTCTGTCAGACCTAGCTAGAGAACACTCCGACGATGGATG--- 954
      508  SerProAlaProAsnProGlyGlnLeuAlaProGlnGlnGlnAlaAspHisLeuGln 527
      953  -----TCCCGACGACGCGCCGCGCTG---CTGGGACATG 918
      528  GlnGlnGlnProProSerProAlaAspProHisSerSerGlyProProMetGlnProAla 547
      917  CAGCAGACGACGAGGGGCTTTAGCTACATTCACACTTGAACACATGATGGGGCT 858
      548  GluSerTrpGlnGln-----IleLeuHisMetSerValLeuGlnHis-----GlyAla 563
      857  GCCCAGCAGACCGTCCGACGG----- 837
      564  AlaAlaGlySerThrProAlaProAlaGlyGlyTyrGlnGlnPheValGlnAlaValys 583
      836  -----AGGGGACGACGCGCCGACGCGGAGACTAGCCTG 804
      584  GlnGlyAlaAlaGlnAspProGlyValProGlyValArgProSerGlyAspProGlyTyr 603
      803  -----AGCTTGCTTACGCGCCACAGGGTT----- 780
      604  LysAlaPheSerSerLeuSerSerAsnGlyIleArgGlyAspThrAlaAlaGly 623
      779  -----GAGAGGACGATCATACCCCATGATTCCTGACATGAGACCA----- 728
      624  ThrAspArgGlyHisGlyGlyTyrLysPro---PheGlnAsnProValProAsnGlnSer 642
      737  -----GCAGTCCCATAGTACGCTGTAACTCACTATTTACATTTGGCATTTATGCA 684
      643  ProSerSerValPro-----LeuPheThrPheGlyLeuAspThr 655
      683  CGTTTGCTTGTCGACATCTGTC-----TGCTGCGA 652
      656  GlnLeuSerProSerProLeuAsnSerAspProProLysSerProProGluCysLeuGly 675
      651  TGTTAGCTTTGTCGACGAGGGTTCGACGATGACATGACATGCTCCGACGATGCACTC 592
      676  LeuGlnLeuGlyLeuLysGly-----GlyAspTrp-----ValLysAlaPro 689
      591  CCTCCACAG-----GCAGCCGACGCTGAGAGATGACTGACGTCGACGCCGAC 541
      690  ProProAlaAspGlnValProLysProPheGlyAspAsp---LeuGlyPheGlyIleValTy 709
      540  CTCGATAGTACGACGACGAGGCTTCATATGTCGTTGTGGATCAGGACCACTT 481
      709  rSerSerLeuThrCys-----HisLeuGlyGly-----HisLeuGlyGlnHis-- 723
      480  GGTTCATCCAGTCTTTGAAAAAGACGACGCTCCGACCTGGGGGACGAGGTGAGAGGTAA 421
      724  ---HisSerGlnGlnGlnGlyGlnSerProIle-----ValAlaSerPr 738
      420  CAGGCGTGGATGTTCTCTTCTTGAGAGCAAGACATTCTCCTAATTGACTGCTCGTA 361
      738  oglyCysGlyCys-----CysTyrAs 745
      360  AGACAGCATACCTCTCTCATACCTCAGACACCCCTCTGACTTCGATGGAGGATCA 301
      745  pasParSerProSerLeuGlySer----- 753
      300  CTATCTATCAACCTCCAGAGAGGCTCTGACATCTCCCTTGAAGGGCTTCCCACTGGC 241

```

```

Db      754  ---LeuSerGlyAlaLeuGlnSerCysPro-----GlnGlyIleProProGlnAl 769
      240  TCACCTGTCACGAGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 181
      769  aasnLeuSerAla-----ProLysThrProSerAsnLeuSerGlyGlnG 785
      180  CAGACATCCACGACCTCAGCCA 159
      785  ylyGlyGlyProGlyHisSerPro 792

RESULT 9
NM33 HUMAN STANDARD; PRT; 1233 AA.
ID NM33 HUMAN
AC 014957;
DT 01-NOV-1997 (Rel. 35, Created)
DT 01-NOV-1997 (Rel. 35, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE Glutamate (NMDA) receptor subunit epsilon 3 precursor (N-methyl
DE D-aspartate receptor subtype 2c) (NR2C) (NMDAR2C).
GN GRIN2C.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Brain;
RX MEDLINE=97189248; PubMed=9037519;
RA Lin J.Y., Boveito S., Carver J.M., Giordano T.;
RT "Cloning of the cDNA for the human NMDA receptor NR2C subunit and its
RT expression in the central nervous system and periphery.";
RL Brain Res. Mol. Brain Res. 43:57-64(1996).
CC - FUNCTION: NMDA RECEPTOR SUBTYPE OF GLUTAMATE-GATED ION CHANNELS
CC POSSESSES HIGH CALCIUM PERMEABILITY AND VOLTAGE-DEPENDENT
CC SENSITIVITY TO MAGNESIUM AND IS MEDIATED BY GLYCINE.
CC - SUBUNIT: HETERODIMER OF AN EPSILON SUBUNIT AND A ZETA SUBUNIT.
CC - SUBCELLULAR LOCATION: Integral membrane protein.
CC - TISSUE SPECIFICITY: MAINLY IN BRAIN WITH PREDOMINANT EXPRESSION IS
CC IN THE CEREBELLUM, ALSO PRESENT IN THE HIPPOCAMPUS, AMYGDALA,
CC CAUDATE NUCLEUS, CORPUS CALLOSUM, SUBTHALAMIC NUCLEI AND THALAMUS.
CC DETECTED IN THE HEART, SKELETAL MUSCLE AND PANCREAS.
CC - SIMILARITY: BELONGS TO THE LIGAND-GATED IONIC CHANNEL FAMILY.
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC
DR EMBL; L76224; AAA88096.1; -
DR HSSP; P19491; IGR2.
DR Genew; HGNC:4587; GRIN2C.
DR MIM; 138254; -
DR InterPro; IPR001320; Ion_glu_receptor.
DR InterPro; IPR001311; SBP/glu_receptor.
DR Pfam; PF00060; lig_chan; 1.
DR ProDom; PD000500; Ion_glu_receptor; 1.
DR SMART; SM00079; PDBE; 1.
KW Receptor; Signal; Transmembrane; Postsynaptic membrane; Calcium;
KW Ionic channel; Magnesium.
FT SIGNAL 1 19
FT CHAIN 20 1233
FT TRANSMEM 554 574
FT TRANSMEM 597 617
FT TRANSMEM 627 647
FT TRANSMEM 815 835
FT SITE 612 612
FT CARBOHYD 70 70
FT CARBOHYD 337 337

```











FT REPEAT 1168 1208 LIN/NOTCH 1.  
 FT REPEAT 1209 1242 LIN/NOTCH 2.  
 FT REPEAT 1243 1282 LIN/NOTCH 3.  
 FT REPEAT 1628 1657 ANK 1.  
 FT REPEAT 1661 1691 ANK 2.  
 FT REPEAT 1695 1724 ANK 3.  
 FT REPEAT 1728 1757 ANK 4.

## Alignment Scores:

Pred. No.: 0.441 Length: 1964  
 Score: 112.50 Matches: 130  
 Percent Similarity: 27.12% Conservative: 36  
 Best Local Similarity: 21.24% Mismatches: 161  
 Query Match: 4.45% Indels: 286  
 DB: 1 Gaps: 34

US-09-989-919-15 (1-1397) x INTCA\_MOUSE (1-1964)

QY 6 TGCACCTGTACCGAGGGGAGATATGACAACTCCA----- 44  
 Db 903 CysHLeaspHrHglYserSerTyRheCySArGcYseProHglYrheHglYlYsLeu 922  
 QY 45 -----CGCAAGCAGCAGTA 59  
 Db 923 CysGlnAspHsMetAsnProCySGluProAsnProCySHiSHiSGlYserThrCySVal 942  
 QY 60 CCGAGTACAGT-----GTATCCACAGACACATCCCGGAGAAAGTACCGCTGC- 112  
 Db 943 ProGlnProSerGlyTyRValCySGlnCySArGcYseProHglYrheHglYlYsLeu 962  
 QY 113 -----TGCCATCTTACCAACACAGGAGCTGCCTC- 143  
 Db 962 eRlYsValLeuGluAlaCySGlnSerGlnProCySHiAsnHISglYThrCys-ThrSer 991  
 QY 144 -----TTTCACTGTTC-ACCTGGCTGAGGCTGT- 171  
 Db 982 ArgProGlnGlyRheHIScYSArGcYseProHglYrheHglYlYsLeuArg-CysGlnGlu 1001  
 QY 172 -----GGATGTCTGTGAGAGCCATGCCAGTGTGGGC- 204  
 Db 1001 YAspValAspGlnCySLeuAspArgProCySHiSerSerGlyThrAlaAlaCySHiSse 1021  
 QY 204 ----- 204  
 Db 1021 rLeuAlaAsnAlaRheTyRcYSGlnCySLeuProGlnHISThrGlnAlaRgCySGluVa 1041  
 QY 205 -----CTTGTGTGTACCAACCA- 224  
 Db 1041 IGluMetAspLeuCySGlnSerGlnProCySserAsnGlyGlySerCySGlnHISThrTh 1061  
 QY 225 -----CCACCTGACAGTGCAGCCAGTGGAGAGCCCTTCCAAAGG-AG 268  
 Db 1061 rGlyProProProGlnRheThrCySHiSYProlYsglyRheGlnGlyProThrCysse 1081  
 QY 269 ATGCAGAGACCTCTCTGAGAGTTGATAGTATGCCCATCGGAAGTAGAGGGGGT 328  
 Db 1081 rHISLeuAlaLeuSerCySgLy-----HISHIScYSHiAsn- 1094  
 QY 329 GCTGAGTGATGAGAGAGGTATACGTGTTCAGGAGCTCAATTAGGAGAGATGT 388  
 Db 1095 -----GlyGlyLeuCy 1098  
 QY 389 CTGCTCTCAGAAAGAAACATCCAGCCCTGTACCTCTCACTC- 434  
 Db 1098 sLeuProSerProGlySerProGlySerProLeuCySArGcYseUserGlyRheGlyG1 1118  
 QY 435 -----TGCCCGCCAGGTGGAGAGTGGGTCTTTTCAAGACTGG 472  
 Db 1118 YProAspCysLeuThrProProAlaProGlnY-----CysG1 1131  
 QY 473 ATGAGCCAGAGTGTCTGTATCC-----AACAAAGCACAATATGAAAGGCTCTGG 526  
 Db 1131 Y-----ProProSerProCyLeuHISAsnGlyThrCySHiGlnGlnThrProG1 1147

QY 527 CTGACCTATCTGAGGGCTGGCTGACACGACTGACTATCTCAGCAGCTGGGCTTGCCTGT 586  
 Db 1147 Y-----LeuGlyAsnPro-- 1151  
 QY 587 GGAGGAGAGTGACTTGCACCTGCGACGACTGCATGTCACTGGAAACCCCTGACAAAGC 646  
 Db 1152 -----GlyRheGlnCysThr-----CysProProAspSerProGlyProArgCy 1166  
 QY 647 TAACTCCGACAGACAGATGTACCGAGCA-----AGTCCAAATATAGC 694  
 Db 1166 sGlnArgProGlnAlaSerGlyCySGlnGlyArgGlyGlyAspGlyThrCys--AspAl 1185  
 QY 695 CAATAGTAAATGATGATTTTACAGCCCTAGCTATGAGAC----- 734  
 Db 1185 aGlyCys-----SerGlyProGlnGlyAspTrpAspGlyGlyAspCysSerLeuG1 1202  
 QY 735 -----TGCTGCTCTTACTTCAGGA 754  
 Db 1202 YValProAspProTrpYsglyCySProProHISserGlnCysTrpLeuLeuPheArgAs 1222  
 QY 755 ATCATG-----GGTATGACTGCT 775  
 Db 1222 rGlyArgCySHiSProGlnCysAspSerGlnGlnCySLeuPheAspGlyTyRAspCys-- 1241  
 QY 776 CTCCAACCTGTG----- 788  
 Db 1242 -GlnLeuProLeuThrCysLeuProAlaTyRAspGlnTyRcYSArGAspHISpHeHISAs 1261  
 QY 789 -----GGCTGAACCAAGCTCAGCTAGTCTCCACCTGGGGGCTGTC 832  
 Db 1261 nGlnHIScYSGlnYsglyCySAsnAsnAlaGluCySGlyTrpAspGlyGlyAspCySAr 1281  
 QY 833 CCTCTCTGGACGGTCCGTGGGCAAGCCCATCACTGTGTTCATATGTTGAGAATGA 892  
 Db 1281 gProGlnGlyGluAspSerGlnGlyArgProSerLeuAla----- 1294  
 QY 893 GGTAAAGCCCTGCTGCTGCTGCTGCAATGCCACAGACGCGGGGCTGCGTGGG 952  
 Db 1295 -----LeuLeuValValLeuArgProPro-----AlaLeuAs 1305  
 QY 953 ACAATCATCGTGAAGTGTCTCTGCACTTACGCTTGACAGCAGAACTTGGCGGGGATG 1012  
 Db 1305 rGlnGlnLeuLeuAlaLeuAlaArgValLeuSerLeuThrLeuArgValGly- 1322  
 QY 1013 CTCAGAGATGTGGGTATCTGTACTGGGAGCGTATCTGCACCTCCGACAGGCGAC 1072  
 Db 1323 -----LeuTrpVal-----ArgYAspSerGlnGlyArgAsnMetVa 1335  
 QY 1073 ACTCCCA-----GGCCAGCCGAGGGGTCA 1096  
 Db 1335 lPheProTyRProGlnThrArgAlaYsglnGlnLeuSerGlyAlaArgAspSerSerse 1355  
 QY 1097 GGGGACAGAGTGCACACTCAGCATGAGCA-----AGACTGGGTC 1138  
 Db 1355 rTrpGlnArgGlnAlaProProThrGlnProLeuGlnYsglnThrGlnSerLeuGlyAl 1375  
 QY 1139 AGGG-----ACGAGTGTGTTGAGCGACGACTGGG 1171  
 Db 1375 aGlyPheValValMetGlyValAspLeuSerArgCySGlyProGlnHISProAlaSe 1395  
 QY 1172 GCGG---GGGTGGGGCGGGGCTTCTGCTCATATTTGCTTCAATGAAGCTCAAGC 1228  
 Db 1395 rArgCySProTrpAspSerGlyLeuLeuLeuArgPheLeu-----AlaAlaMetAl 1412  
 QY 1229 AGCCAAACCAAGCTTCCCTTCTC 1256  
 Db 1412 aAlaValGlyAlaLeuGlnProLeuLeu 1421

RESULT 12  
 VGP3\_EBV STANDARD; PRT; 907 AA.  
 ID VGP3\_EBV  
 AC P03200; P03201;

DT 21-JUL-1986 (Rel. 01, Created)  
 DT 21-JUL-1986 (Rel. 01, last sequence update)  
 DT 16-OCT-2001 (Rel. 40, last annotation update)  
 DE Envelope glycoprotein GP340 (Membrane antigen) (MA) [Contains:  
 GN Glycoprotein GP220].  
 GN BLRP1.  
 OS Epstein-Barr virus (strain B95-8) (Human herpesvirus 4).  
 OC Viruses; dsDNA viruses, no RNA stage; Herpesviridae;  
 OC Gammaherpesvirinae; Lymphocryptovirus.  
 NC NCB1\_TaxID=10377;  
 RN [1].  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE=84270667; PubMed=6087149;  
 RA Baer R., Bankier A.T., Biggin M.D., Deininger P.L., Farrell P.J.,  
 RA Gibson T.J., Hatfull G., Hudson G.S., Satchwell S.C., Seguin C.,  
 RA Tuffnell P.S., Barrett B.G.;  
 RT "DNA sequence and expression of the B95-8 Epstein-Barr virus genome";  
 RL Nature 310:207-211(1984).  
 CC -1- FUNCTION: RESPONSIBLE FOR EBV BINDING TO THE CR2 RECEPTOR ON HUMAN  
 CC B-CELLS.  
 CC -1- SUBCELLULAR LOCATION: MOST ABUNDANT COMPONENT OF THE VIRAL  
 CC ENVELOPE.  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC  
 DR EMBL; V01555; CAA24854.1; -  
 DR PIR; A03762; Q0BE21.  
 DR PIR; S33008; S33008.  
 KM Membrane; Glycoprotein; Antigen; Late protein; Alternative splicing.  
 FT CARBOHYD 47 47 N-LINKED (GLCNAC...) (POTENTIAL).  
 FT CARBOHYD 87 87 N-LINKED (GLCNAC...) (POTENTIAL).  
 FT CARBOHYD 114 114 N-LINKED (GLCNAC...) (POTENTIAL).  
 FT CARBOHYD 166 166 N-LINKED (GLCNAC...) (POTENTIAL).  
 FT CARBOHYD 169 169 N-LINKED (GLCNAC...) (POTENTIAL).  
 FT CARBOHYD 195 195 N-LINKED (GLCNAC...) (POTENTIAL).  
 FT CARBOHYD 229 229 N-LINKED (GLCNAC...) (POTENTIAL).  
 FT CARBOHYD 277 277 N-LINKED (GLCNAC...) (POTENTIAL).  
 FT CARBOHYD 318 318 N-LINKED (GLCNAC...) (POTENTIAL).  
 FT CARBOHYD 328 328 N-LINKED (GLCNAC...) (POTENTIAL).  
 FT CARBOHYD 345 345 N-LINKED (GLCNAC...) (POTENTIAL).  
 FT CARBOHYD 356 356 N-LINKED (GLCNAC...) (POTENTIAL).  
 FT CARBOHYD 378 378 N-LINKED (GLCNAC...) (POTENTIAL).  
 FT CARBOHYD 386 386 N-LINKED (GLCNAC...) (POTENTIAL).  
 FT CARBOHYD 411 411 N-LINKED (GLCNAC...) (POTENTIAL).  
 FT CARBOHYD 435 435 N-LINKED (GLCNAC...) (POTENTIAL).  
 FT CARBOHYD 443 443 N-LINKED (GLCNAC...) (POTENTIAL).  
 FT CARBOHYD 457 457 N-LINKED (GLCNAC...) (POTENTIAL).  
 FT CARBOHYD 497 497 N-LINKED (GLCNAC...) (POTENTIAL).  
 FT CARBOHYD 519 519 N-LINKED (GLCNAC...) (POTENTIAL).  
 FT CARBOHYD 533 533 N-LINKED (GLCNAC...) (POTENTIAL).  
 FT CARBOHYD 547 547 N-LINKED (GLCNAC...) (POTENTIAL).  
 FT CARBOHYD 568 568 N-LINKED (GLCNAC...) (POTENTIAL).  
 FT CARBOHYD 589 589 N-LINKED (GLCNAC...) (POTENTIAL).  
 FT CARBOHYD 610 610 N-LINKED (GLCNAC...) (POTENTIAL).  
 FT CARBOHYD 624 624 N-LINKED (GLCNAC...) (POTENTIAL).  
 FT CARBOHYD 627 627 N-LINKED (GLCNAC...) (POTENTIAL).  
 FT CARBOHYD 645 645 N-LINKED (GLCNAC...) (POTENTIAL).  
 FT CARBOHYD 656 656 N-LINKED (GLCNAC...) (POTENTIAL).  
 FT CARBOHYD 683 683 N-LINKED (GLCNAC...) (POTENTIAL).  
 FT CARBOHYD 701 701 N-LINKED (GLCNAC...) (POTENTIAL).  
 FT CARBOHYD 735 735 N-LINKED (GLCNAC...) (POTENTIAL).  
 FT CARBOHYD 746 746 N-LINKED (GLCNAC...) (POTENTIAL).  
 FT CARBOHYD 755 755 N-LINKED (GLCNAC...) (POTENTIAL).  
 FT CARBOHYD 780 780 N-LINKED (GLCNAC...) (POTENTIAL).  
 FT CARBOHYD 815 815 N-LINKED (GLCNAC...) (POTENTIAL).  
 FT CARBOHYD 858 858 N-LINKED (GLCNAC...) (POTENTIAL).

FT CARBOHYD 888 888 N-LINKED (GLCNAC...) (POTENTIAL).  
 FT VARSEPLIC 502 698 MISSING (IN GP220).  
 SQ SEQUENCE 907 AA; 34431 MW; 0750141CBAC52C9 CRC64;  
 Alignment Scores:  
 Pred. No.: 0.439  
 Score: 112.00  
 Percent Similarity: 32.53%  
 Best Local Similarity: 21.69%  
 Query Match: 4.38%  
 DB: 1  
 Gaps: 18  
 US-09-989-919-15 (1-1397) x VGP3\_EBV (1-907)  
 QY 1182 CCCACCCCGCCCGCCAGT---CCTGGCTCAAAACACAC--- 1147  
 DB 442 ProAsnThrThrThrglyLeuProSerSerThrHisValProThrAsnLeuThrAlaPro 461  
 QY 1146 TGCTCCCTGACCCCGACCTTGGCTCATGTGAGTGTGACACCTCCCTGACCCCTGG 1087  
 DB 462 AlaSerThrGlyProThrValSerThrAlaAspVal-----ThrSerPro----- 476  
 QY 1086 GCTGCGCTGGAGTGTCCCTGTGCGAGGTGACAGATGCTCCCGAGTACAGATCA 1027  
 DB 477 -----ThrProAlaGlyThrThrSerGlyAlaSerProValThrProSer 491  
 QY 1026 CCACATCTGAGCATGCCCGCCGCAAGTCTCCTGTCGACACTAGACAGAGAACT 967  
 DB 492 ProSerProThrAspAsnGlyThrGluSerLysAlaProAspSerThrSerThrSer 511  
 QY 966 CCACATGATGATGTCCCGACGACGCCCGCTGCTGTCGATGTCAGACAGACGACGA 907  
 DB 512 Pro-----ValThrThrProThrProAsnAlaThrSerProThrProAlaVal 527  
 QY 906 GCAGGGGCTTTAGCTACATTCTACACTATTGAACAGATGAGGGCTGCCACGGAAC 847  
 DB 528 Thr----- 528-  
 QY 846 CGTCCAGGAGGAGGACAGACCCCGAGTGGGAGACTAGCTGAGCTTCTTACAGCCGA 787  
 DB 529 -----ThrPro 530  
 QY 786 CAGGTTGAGAGGAGCATATACCCCATGAT- 754  
 DB 531 ThrProAsnAlaThrSerProThrProAlaValThrThrProThrProAsnAlaThrSer 550  
 QY 753 CTGAGCTAGAGACGACAGTCCCATAGCTAGGCTGTAACATTTTAACATTTGG 694  
 DB 551 ProThrLeuGlyLysThrSerProThrSerAlaValThrThrProThrProAsn----- 568  
 QY 693 CATTAATTGACGTTTCTCTGTGACATCTGTCTGCGAGATGTTAGCTTGTCTGCGAG 634  
 DB 569 -----AlaThrSerProThrLeuGly-LysThr-----SerPro 579  
 QY 633 GGGTCCAGAGGATGATGAGTGTCTCCAGTCAATCTACCTCCACAGGACAG- 578  
 DB 579 cthrsrAlaValThrThrProThrProAsnAlaThr-serProThrLeuGlyLysThr 599  
 QY 577 -----CCACGCTGAGAGATGATGCTGCTGACG 547  
 DB 599 exProThrSerAlaValThrThrProThrProAsnAlaThrGlyProThrValGlyGlu 619  
 QY 546 CGAGCCCTCAG----- 536  
 DB 619 hsrerProGlnAlaAsnAlaThrAsnHisThrLeuGlyLysThrSerProThrProVal 639  
 QY 535 --ATAGTACAGCAGAGGCTTCAATATGTTGTTGGATGAGGACCACTTGGC 478  
 DB 639 althrserGlnProLysAsnAlaThrSerAlaVal-----ThrThrGlyG 654  
 QY 477 TCCATCCAGCTTTGAAAAAGACC-----AGTCCGACCTGAGGGGACAGAGT 430  
 DB 654 InhlAsnIleThrSerSerSerThrSerSerMetSerLeuThrProSer----- 670

OY	429	GAGAGGTACACGGCTGATGTTTCTCTTTTGAGAGCGAACAACATTCTCCATAATTGA	370
Db	671	-----SeraspProgluthrLeuSerProserThrS	681
OY	369	CTGCCTTGAAACACAGTATACCCTCTCTTCATCACTCAGACGCCCTCTGACTTCGCAT	310
Db	661	:::::Thr 697	
OY	309	GGGGGATACATCTATCATCAACCTCCAGAGAGTCTCGCATCTCCCTTTGAAAGGCTTCT	250
Db	698	GlyglyAlu-----AsnIle-ThrGluValThrProAlaSerIle	710
OY	249	CCCACTGGCTCACCTGTCACAGGTGGTGTGGTTGGACCAACAAGGCCCCGACACTGGACA	190
Db	711	-----SerThrHisH1	714
OY	189	TGGCTCTCAGACACATCCACAGCCCTCAGCCAGGTGAACACTGAAGAAGGACACTCCCG	130
Db	714	SvalSerThrSerSerProAlaProArgProGlyThrThr-----SerGlnAlaIse	731
OY	129	TGGTGTAGATGATGCCAGCAGCGGTAGTCTTCTCGG	93
Db	731	rGlProGlyanserSerThrSerThrLysProGly	743
RESULT 13			
POLS_RUBVM	STANDARD;	PRT;	992 AA.
ID	POLS_RUBVM		
AC	P08563;		
DT	01-AUG-1988 (Rel. 08, Created)		
DT	01-AUG-1988 (Rel. 08, Last sequence update)		
DT	16-OCT-2001 (Rel. 40, Last annotation update)		
DE	Structural polyprotein [Contains: Nucleocapsid protein C; Membrane glycoprotein E2; Membrane glycoprotein E1].		
OS	Rubella virus (Strain M33).		
OC	Viruses; ssRNA positive-strand viruses; no DNA stage; Togaviridae; Rubivirus.		
OX	NCBI_TaxID=11043;		
RN	[1]		
RP	SEQUENCE FROM N.A.		
RX	MEDLINE=87174825; PubMed=3562245;		
RA	ClaRke D.M., Loo T.W., Hui I., Chong P., Giliam S.;		
RT	"Nucleotide sequence and in vitro expression of rubella virus 24S subgenomic messenger RNA encoding the structural proteins E1, E2 and C."		
RL	Nucleic Acids Res. 15:3041-3057(1987).		
CC	- FUNCTION: GLYCOPROTEIN E1 CONTAINS THE VIRAL HEMAGGLUTININ ACTIVITY.		
CC	- SUBUNIT: MULTIPLE COPIES OF THE C PROTEIN COMPRISE THE NUCLEOCAPSID.		
CC	- DISASE: RUBELLA INFECTION DURING THE FIRST TRIMESTER OF PREGNANCY RESULTS IN DEATH OR MULTISYSTEM BIRTH DEFECTS & MICROCEPHALY.		
CC	This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its contents in no way can be modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See http://www.isb-sdb.ch/announce/ or send an email to license@isb-sdb.ch).		
CC	EMBL; X05259; CNA28880.1; -		
DR	PIR; A27505; GNWVR3.		
KM	Polyprotein; Nucleocapsid; Glycoprotein; Transmembrane; Signal.		
FT	CHAIN	1	229
FT	CHAIN	300	559
FT	SIGNAL	560	580
FT	CHAIN	581	992
FT	TRANSMEM	277	297
FT	TRANSMEM	515	531
FT	TRANSMEM	533	553
FT	CARBOHYD	352	352
FT	CARBOHYD	370	370
MEMBRANE GLYCOPROTEIN E1. MEMBRANE GLYCOPROTEIN E2. MEMBRANE GLYCOPROTEIN E1. POTENTIAL. POTENTIAL. POTENTIAL. N-LINKED (GLCNAC... ) (POTENTIAL). N-LINKED (GLCNAC... ) (POTENTIAL).			

FT	CARBOHYD	656	656	N-LINKED (GLCNAC. . .)	(POTENTIAL)
FT	CARBOHYD	757	757	N-LINKED (GLCNAC. . .)	(POTENTIAL)
FT	CARBOHYD	789	789	N-LINKED (GLCNAC. . .)	(POTENTIAL)
SQ	SEQUENCE	992 AA;	106905 MW;	D268688C9E78CF1E	CRC64;

Alignment Scores:

Query Match:	4.38%	Indels:	15
Best Local Similarity:	22.99%	Mismatches:	12
Percent Similarity:	31.99%	Conservative:	38
Score:	112.00	Matches:	97
Refid. NO.:	0.47%	Menguan:	9
DB:	1	Gaps:	22

US-09-989-919-15 (1-1397) x POLS\_RUBVM (1-992)

QY	1184	GGCCACACCCCGCCAGGCTCTGGCTCAAC	-----ACACCTGCTCCGACCC	113
Db	76	AlAProPProPProGluGluArgGlnuGserGIntnProAlaProLysPro	-----	95
QY	1133	CAGTCTTGAGCTCATGCTAGAGGTGGACCTCTGCCCCCTGACCCCTGGGGCTGGCCTGGGAG	107	
Db	96	SerArg-----AlaPro-----ProGlnGlnProGln-----	104	
QY	1073	TGTCCCTGTGGGAGGTCAAGATAGCTCCCGCAGTAGAG-----	-AAT 1023	
Db	105	-----ProProArgMetGlnThrGlyuArgGlyuGlyser	115	
QY	1028	CACCCACATCCTGGAGCATCCCGCCCAAGTCTCTGTCCAGACCTTAAGCTAGAGA---	972	
Db	116	AlAProArgProGluLeuGlyProProThrAsnProPheGlnAlaAlaValAlaArgGly	135	
QY	971	-----ACATCTCAGATGATGATGCCACGACGCCCCCAGCGCTGTG---	926	
Db	136	LeuArgProProLeuGlnAsnPro---AspThrGluAlaProThrGlnAlaCysValTh	154	
QY	935	---GGCATGTGCACGACAGCAGCAGGAGGCGCTTAAGTCAATTTCACTATTGAACAC	870	
		:::	:::	
Db	154	rSerTrpLeuTrpSerGluGluGlyuAlaValaPheTykArgValaAspLeuHisrPhe11	174	
QY	869	AGTATGGGGCGGCCACGGAACCGTCCAGGGA-----	836	
Db	174	eAsnLeuGluThrProProLeuAspGlnAspGlnArgTrpAspProAlaLeuMetTykAs	194	
QY	835	-----GGGACACAGCCCCAGTGGGAGACTAGCAGCTTGCTTAACGCCACAG	783	
		:::	:::	
Db	194	nProCysGlyProGlnProPro---AlaHisValAlaArgAlaTykAsnGlnProAlaG1	213	
QY	782	GTTGAGAGGCGAGTCAATACCCCATGATCTCCGAGTACAGGACCGACAGCTGCCATAGTA	723	
Db	213	AspValaArg-----GlyValaTrpGlyuGlyuGlnuArgThrTykAla-	227	
QY	722	GGCTGGTAACTCAATTTTAACAT-----TGGCATTAATTCACAGTTTGTC	675	
Db	228	-----GluGlnAspPheArgValaGlyuThArgTrpHisArgLeuLeuArgMetP	245	
QY	674	TGCTACATCTGTCTGTCTGGAGTGTAGCTTTGCTGACAGGCGTTCCAGAGTACATGC	615	
Db	245	roVal-----ArgGlyLeuAspGlyuAsp-----	252	
QY	614	AGTGTGCAGTGCAGAGTCACTCCCTCCACAGCAAGCCCACTGCTGAGATAG-TCAG	556	
Db	253	-----ThrAlaProLeuProProGlnHisThrThrGlnArgGlnuThrArgSerA	269	
QY	555	CTGGTCAAGCG---AGCCTCAATAGTCAAGCCAGAGGCTTTCACATATGGGTGTTGT	499	
Db	269	lArgHisAspProTrpArgGlnArgPheGlyuAlaProGlnAlaPheLeuAlaGlyLeuLeu	289	
QY	498	TGGAGTACAGGAGCACTTGCTGCTCACTCCAGTCTGTAAGAAAGACAGCTCCGACCGGGG	439	
		:::	:::	
Db	289	eAlaAlaValaAla-ValaGlyThrAlaArgAlaGlyuGlnProArgAlaAspMet---	307	
QY	438	GGCAGAGGTGAGAGGTAAACAGGCGCTGATGTTTCTCTTTGAGGCAAGACCATTTGCC	379	

Db 307 ----- 307  
 QY 378 CTAATTGACTGCTTGAAGACAGTATACCTCTCTCTCATCAGACACCCCTG 319  
 Db 308 ----- AlalalProPome 312  
 QY 318 ACTTCGATGGGGATCACTATCTATCACTCCAGAGAGCTCCCTCCCTTGA 259  
 Db 313 ----- Pro-ProGlnPro 316  
 QY 258 AGGCTTCTCCCACTGCTCACTCTCCAGTGTCTGTGTGTGACCAAGGCCCA 199  
 Db 317 ----- ProArgAlaHisGlyGlnHisTyrGlnHisHisGlnLeuProH 333  
 QY 198 CACTGGCATGAGCTCTCAAGACATCCAGCTCAGCAGTTGAACACTGAAGAG 139  
 Db 333 eleuGlyHisArgGlyHisHis ----- GlyGly 342  
 QY 138 CAGCTCCCTGCTGTGATGATGCGCAGCGGACT ----- CT 100  
 Db 342 yThrLeuArgValGlyGlnHisHisTyrGlnHisHisGlnLeuProGlyHisTyr 362  
 QY 99 TCTGGGGATGTGCTGTCTGTGGA ----- TACACTGTGACTCGTACTGC 54  
 Db 362 uGlnGlyGlyTyrGlyCysTyrHisLeuSerAspTyrHisGlnGlyThrHisValCys 381  
 RESULT 14  
 TEGU\_EBV STANDARD; PRT; 3149 AA.  
 ID TEGU\_EBV  
 AC P03186;  
 DT 21-JUL-1986 (Rel. 01, Created)  
 DT 21-JUL-1986 (Rel. 01, Last sequence update)  
 DT 01-APR-1993 (Rel. 25, Last annotation update)  
 DE Large tegument protein.  
 GN BPLF1  
 OS Epstein-Barr virus (strain B95-8) (Human herpesvirus 4).  
 OC Viruses; dsDNA viruses, no RNA stage; Herpesviridae;  
 OC Gammaherpesvirinae; Lymphocryptovirus.  
 NC NCBI\_TaxID=103177;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE=84270667; PubMed=6087149;  
 RA Baer R., Bankier A.T., Biggin M.D., Deininger P.L., Farrell P.J.,  
 RA Gibson T.J., Hatfull G., Hudson G.S., Satchwell S.C., Seglin C.,  
 RA Tufnell P.S., Barrett B.G.;  
 RT "DNA sequence and expression of the B95-8 Epstein-Barr virus genome";  
 RL Nature 310:207-211(1984).  
 CC -1- FUNCTION: TEGUMENT PROTEIN.  
 CC -1- SIMILARITY: BELONGS TO FAMILY THAT GROUPS TOGETHER HSV-1 UL36,  
 CC EHv-1 24, EBV BPLF1, HVS-1 64, VZV 22, AND HCMV UL48.  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation-  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See http://www.isb-sib.ch/announce/  
 CC or send an email to license@isb-sib.ch).  
 CC  
 DR EMBL; VO1555; CAA24839.1;  
 DR PIR; A03747; O0B8.  
 DR PIR; S32993; S32993.  
 SO SEQUENCE 3149 AA; 337954 MW; 3DD0C57658713DB8 CRC64;  
 Alignment Scores:  
 Pred. No.: 0.51 Length: 3149  
 Score: 112.00 Matches: 81  
 Percent Similarity: 34.38 Conservative: 40  
 Best Local Similarity: 23.01 Mismatches: 116  
 Query Match: 4.38 Indels: 16  
 Gaps: 16  
 US-09-989-919-15 (1-1397) x TEGU\_EBV (1-3149)

QY 1193 AAGCCCCGGCCCCCACCACCCGCCCG-----GTCTGCTCAACACACACTGCTCC 1140  
 Db 2624 GluAlaSerGlyProProSerProGlnSerProLeuLeuGlnValAlaProGlyArg 2643  
 QY 1139 TGAACCCAGCTTGGCTCACTGCTGAGTGTGACCT-----CTGCCCTGACCCCTGG 1086  
 Db 2644 ---ProArgAspTyrLeuSerProThrSerSerProArgAspValThrValThrProGly 2662  
 QY 1085 CTGGCTGGAGATGTCCTCTGTCGAGAGTCAAGATTAACCTCCAGCTACAGATC 1026  
 Db 2663 LeuAlaAlaProIleThrLeuProGly-----ProArgLeuMetAlaArg 2677  
 QY 1025 CCACATCTCGAGACATCCCC-CGCCAAGTCTCTGTCAGACCTAAGTGAAGACACT 967  
 Db 2678 ProTyrPheGlyAlaGlnThrArgAlaSerGlnSerProHisArgSerProGlySerSer 2697  
 QY 966 CCACGA---TGG-----ATTGTCGCCAGCAGCCCA-----CCG 934  
 Db 2698 ProArgProTyrProHisAspSerLeuGlnLeuLeuProGlnProAlaProGlnGlnPro 2717  
 QY 933 CTGCTGTGGCATGTGCACAGCAGCAGAGGGCTTTAGTACATTCACACTATTGA 874  
 Db 2718 ProSerSerProTyrAlaSerGlnGlnGlyProIleValTyrThrLeuSer----- 2734  
 QY 873 ACACAGTATGGGGCTGCGCCACGAAACGCTCCAGGAGGAGGACAGCCCGAGTGGGA 814  
 Db 2735 -----ProHiserThrProSer----- 2740  
 QY 813 GACTAGCTGAGCTTGCTTACAGCCACAGGGTTGGAGAGCATGATACCCCATGATT 754  
 Db 2741 -----ThrAlaSerGlySerGlnLysHisThrIleGlnIle 2753  
 QY 753 CTTGACTAGTGAAGCCAGCAGAGTCCCATAGCTAGCTGTGTAACCTATTAACTTTG 694  
 Db 2754 ProGlyLeuValProSer----- 2759  
 QY 693 CATTAATGACGTTGTCTGTGATCATATGTCGTCTGTGAGATGTAGCTTTGTCTGAG 634  
 Db 2759 ----- 2759  
 QY 633 GGGTCCAGAGTGAATGACAGTGTCCAGTCAGATCACTCCCTCCAGAGCA----- 580  
 Db 2760 -----GlnLysProSerTyrProProSerAlaProTyr 2770  
 QY 579 -----AGCCAGCTGCTGAGGATAGTACGTGCTGAC 547  
 Db 2771 LysProGlyGlnSerThrGlyGlyLeuAlaProThrProSerAlaAlaSerLeuThrThr 2790  
 QY 546 CGAGCCCTCAGATAGGTAGCCAGAGGCTTCACATATGATGATCTTGGGATCAGGA 487  
 Db 2791 PheGlyLeu-----GlnProGlnAspThrGlnAlaSerSerGlnAspPro 2805  
 QY 486 CCACTTGGCTCCATCCATCTTGAAGAAAGACAGCTGCCAGCTGGGGGGAGAGTGA 427  
 Db 2806 ProTyrGly-HisSerIleMetGlnArgGlnLysGlnGly-----ArgGlnG 2824  
 QY 426 AGCTAACAGGCTGATGTTCTCTTTCTTGAGGAGCAAGCATTCCTTAATTGACTG 367  
 Db 2824 LysAlaAlaGlnIleArgProSer-----AlaThrArgLeuProThrAlaValG 2840  
 QY 366 CCTTGAGACAGATATACCTCTCTC-----ATACCTCAGACACCCCTCTG 319  
 Db 2840 LysLeuArgProArgAlaProVal-ValAlaAlaGlyAlaAlaAlaSerIleThrProAla 2859  
 QY 318 ACTTCGATGGGGATCACTATCT-ATCACTCCAGAGAGAGTCTCCATCTCCCT 262  
 Db 2860 PheAspProGlyGlnAlaProSerGlyPheProIleProGlnAlaProAlaLeuGlySer 2879  
 QY 261 GGAAGGCTTCTCCACCTGGCTACCTGTC 232  
 Db 2880 GlyLeuAlaAlaProAlaHisThrProVal 2889

RESULT 15  
SFPQ\_HUMAN STANDARD; PRT; 707 AA.

AC P23246; P30808;  
DT 01-NOV-1991 (Rel. 20, Created)  
DT 01-OCT-1996 (Rel. 34, Last sequence update)  
DT 15-JUN-2002 (Rel. 41, Last annotation update)  
DE Splicing factor, proline- and glutamine-rich (Polypyrimidine tract-  
binding protein-associated splicing factor) (PTB-associated splicing  
factor) (PSF) (DNA-binding P52/Plt00 complex, 100 kDa subunit).  
GN SFPQ OR PSF.  
OS Homo sapiens (Human).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
OC Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.  
OX NCBI\_Taxid=9606;  
RN [1]  
RP SEQUENCE FROM N.A., PARTIAL SEQUENCE, AND ALTERNATIVE SPLICING.  
RC TISSUE=Fetal brain;  
RA MEDLINE=3194059; PubMed=8449401;  
RT Patton J.G., Porro E.B., Galceran J., Tempst P., Nadal-Ginard B.;  
RT "Cloning and characterization of PSF, a novel pre-mRNA splicing  
factor.";  
RT Genes Dev. 7:393-406(1993).  
RL [2]  
RP SEQUENCE OF 312-707 FROM N.A.  
RC TISSUE=Fetal skeletal muscle;  
RA MEDLINE=90091812; PubMed=2480877;  
RX Gower H.J., Moore S.E., Dickson G., Elsom V.L., Nayak R., Walsh F.S.;  
RT "Cloning and characterization of a myoblast cell surface antigen  
defined by 24.1D5 monoclonal antibody.";  
RT Development 105:723-731(1989).  
RL [3]  
RP SEQUENCE OF 48-68 AND 213-246.  
RA MEDLINE=93176127; PubMed=8439294;  
RX Zhang W.-W., Zhang L.-X., Busch R.K., Farres J., Busch H.;  
RT "Purification and characterization of a DNA-binding heterodimer of 52  
and 100 kDa from HeLa cells.";  
RT Biochem. J. 290:267-272(1993).  
RL [4]  
RP FUNCTION: ESSENTIAL PRE-MRNA SPLICING FACTOR REQUIRED EARLY IN  
SPLICOSOME FORMATION. BINDS TO THE MAMMALIAN POLYPYRIMIDINE  
TRACTS. FORMS A COMPLEX WITH THE POLYPYRIMIDINE TRACT-BINDING  
PROTEIN (PTB). SEEMS TO ALSO BIND DNA.  
CC -1 SUBUNIT: HETEROTETRAMER OF TWO 52 KDA AND TWO 100 KDA SUBUNITS.  
CC -1 SUBCELLULAR LOCATION: Nuclear.  
CC -1 ALTERNATIVE PRODUCTS: AT LEAST 2 ISOFORMS, A LONG FORM (SHOWN  
HERE) AND A SHORT FORM, ARE PRODUCED BY ALTERNATIVE SPLICING.  
CC -1 SIMILARITY: CONTAINS 2 RNA RECOGNITION MOTIFS (RRM).  
CC -1 CAUTION: WAS ORIGINALLY (REF.2) THOUGHT TO BE MYOBLAST CELL  
SURFACE ANTIGEN 24.1D5 AND A POSSIBLE MEMBRANE-BOUND PROTEIN  
ECTOKINASE.  
CC CC  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC CC  
DR EMBL; X70944; CAAS0283.1; -  
DR EMBL; X16850; CAAS4747.1; -  
DR PIR; A43557; A43557.  
DR PIR; S29770; S29770.  
DR HSSP; P11940; 1CVT.  
DR SWISS-2DPAGE; P23246; HUMAN.  
DR Genew; HGNC:10774; SFPQ.  
DR MIM; 605199; -  
DR InterPro; IPR000504; RNA\_rec\_moc.  
DR Pfam; PF000076; rrm; 2.  
DR SMART; SM00360; RRM; 2.  
DR PROSITE; PS50102; RRM; 2.  
DR PROSITE; PS00030; RRM\_RNP\_1; 1.  
KW Nuclear protein; RNA-binding; DNA-binding; mRNA splicing; Repeat;

KM Alternative splicing.

FT	DOMAIN	297	359	RNA-BINDING (RRM) 1.
FT <td>DOMAIN</td> <td>371</td> <td>452</td> <td>RNA-BINDING (RRM) 2.</td>	DOMAIN	371	452	RNA-BINDING (RRM) 2.
FT <td>DOMAIN</td> <td>9</td> <td>27</td> <td>3 X 3 AA REPEATS OF R-G-G.</td>	DOMAIN	9	27	3 X 3 AA REPEATS OF R-G-G.
FT <td>REPEAT</td> <td>9</td> <td>11</td> <td>1.</td>	REPEAT	9	11	1.
FT <td>REPEAT</td> <td>19</td> <td>21</td> <td>2.</td>	REPEAT	19	21	2.
FT <td>REPEAT</td> <td>25</td> <td>27</td> <td>3.</td>	REPEAT	25	27	3.
FT <td>DOMAIN</td> <td>10</td> <td>266</td> <td>GLN/GLU/PRO-RICH.</td>	DOMAIN	10	266	GLN/GLU/PRO-RICH.
FT <td>DOMAIN</td> <td>10</td> <td>15</td> <td>POLY-GLY.</td>	DOMAIN	10	15	POLY-GLY.
FT <td>DOMAIN</td> <td>20</td> <td>27</td> <td>POLY-GLY.</td>	DOMAIN	20	27	POLY-GLY.
FT <td>DOMAIN</td> <td>56</td> <td>65</td> <td>POLY-PRO.</td>	DOMAIN	56	65	POLY-PRO.
FT <td>DOMAIN</td> <td>67</td> <td>71</td> <td>POLY-GLN.</td>	DOMAIN	67	71	POLY-GLN.
FT <td>DOMAIN</td> <td>95</td> <td>98</td> <td>POLY-GLN.</td>	DOMAIN	95	98	POLY-GLN.
FT <td>DOMAIN</td> <td>99</td> <td>103</td> <td>POLY-PRO.</td>	DOMAIN	99	103	POLY-PRO.
FT <td>DOMAIN</td> <td>184</td> <td>188</td> <td>POLY-ARG.</td>	DOMAIN	184	188	POLY-ARG.
FT <td>DOMAIN</td> <td>571</td> <td>574</td> <td>POLY-GLY.</td>	DOMAIN	571	574	POLY-GLY.
FT <td>DOMAIN</td> <td>613</td> <td>616</td> <td>POLY-GLY.</td>	DOMAIN	613	616	POLY-GLY.
FT <td>VARSPIC</td> <td>663</td> <td>707</td> <td>RTERRGGGAGPVGGGGPRGMRGPRGAGYGRGREGG ENKRRF -&gt; VRMIDVG (IN SHORT ISOFORM). G -&gt; R (IN REF. 3). 6D8D5EA95E235847 CRC64;</td>	VARSPIC	663	707	RTERRGGGAGPVGGGGPRGMRGPRGAGYGRGREGG ENKRRF -> VRMIDVG (IN SHORT ISOFORM). G -> R (IN REF. 3). 6D8D5EA95E235847 CRC64;
FT <td>CONFLICT</td> <td>243</td> <td>243</td> <td></td>	CONFLICT	243	243	
FT <td>SEQUENCE</td> <td>707 AA;</td> <td>76149 NM;</td> <td></td>	SEQUENCE	707 AA;	76149 NM;	

Alignment Scores:

Pred. No.:	0-51	Length:	707
Score:	111.00	Matches:	49
Percent Similarity:	36.758	Conservative:	12
Best Local Similarity:	29.528	Mismatches:	70
Query Match:	4.348	Indels:	35
DB:	1	Gaps:	7

US-09-989-919-15 (1-1397) x SFPQ\_HUMAN (1-707)

QY	1192	AGGCCCGGCGCCACCCCGGCGGCTGCTCAACACACACCTGCTCCCTGACCCC	1133
DB	86	GlnProProProHisProGlnInProHisGlnGlnGlnInProPro-----ProProPro	103
QY	1132	AGCTTGTGCTATGCTGAGGTGACACCTTCGCTGACCCCTGACCCCTGGGCTGGGAGT	1073
DB	104	GlnAspSerSer-LysProValValaLagInGlyProGlyProAlaProGlyValGlySe	123
QY	1072	GTCCTGTCGGGAGGTGACAGATAGCCTCCCGAGGTACAGATACCCACATCTGGAG	1013
DB	123	ralaProProAlaSerSer-----SerAlaProProAlaThrProProThrSerGlyAl	141
QY	1012	CATCCCGCGCAAGCTCTGTCAGACCTAAGCTGAGAGAACCTCCACGATGATTGT	953
DB	141	aproProGly-----SerGlyProGly-----	148
QY	952	CCCCAGCGACCCCGCCACCGCGCTGTCGATGTCAGACGACGACGACGAGGGCTTTAGC	893
DB	149	ProThrProThrProProProAlaAlaValThrSerAlaProProGlyAlaProProProTh	168
QY	892	TACATTTCTACACTATTGAACAGATGATGGGCTGCCACCGAACCGTCCAGAGAGG	833
DB	168	ProProSerSer-----GlyValProThrThrProProGlnAlaG	182
QY	832	GCACAGCCCCCAG-----TGGGAGACTAGCCTGAGCTGCTTA	794
DB	182	YgLYProProProProProAlaAlaValProGlyProGlyProGlyProGlyGlnGlyPr	202
QY	793	CAGCCCGACGGGTGGAGGAGGACGATACCCCGCATGATCTCT	746
DB	202	oGlyProGlyGlyProGlyGlyGlyLysMetProGlyGlyProGlyGlyGlyPr	222
QY	745	AGGAGCCAGCAGTCC	730
DB	222	oGlyLeuSerThrPro	227

Search completed: June 24, 2003, 13:53:37  
Job time : 55 secs